THE ASTROPHYSICAL JOURNAL

AN INTERNATIONAL REVIEW OF SPECTROSCOPY AND ASTRONOMICAL PHYSICS

VOLUME 91 MARCH 1940 NUMBER 2

COSMOLOGICAL THEORIES[1]

E. A. MILNE

ABSTRACT

The present report is mainly concerned with a careful analysis of the fundamental ideas, methods, and postulates which underlie current approaches to the cosmological problem. In particular, the methods of general relativity and the author's kinematical relativity are contrasted. In the later method all the fundamental notions ("uniform time," "rigid length scale") are carefully scrutinized and emphasis is laid on the methods by which two arbitrary observers can set up "congruent" clocks and distance measures by dispatching and receiving light-signals. A *linear equivalence*, defined as a class of observers possessing congruent clocks, is then analyzed. This leads to the consideration of the two scales of time, t and $\tau = t_0 \log t/t_0 + t_0$, such that a uniform motion equivalence in the scale of t becomes one at relative rest in the scale of τ. The physical significance of the two scales of time are then further discussed. The methods by which the dynamical equations of motion are to be set up by projecting a "free" test particle in the presence of a density distribution and specified equivalence are then considered. The fundamental problem here is the determination of the motion of a free particle for an equivalence satisfying the *cosmological principle*, i.e., the case when the equivalence is such that the statistical description of the whole equivalence from any one member of the equivalence coincides with that from any other member. The solution of the cosmological problem which the consideration of this problem leads to is then discussed; particular attention is here given to the degree of arbitrariness which is left at each stage of the solution of the problem and how the arbitrariness is then later removed. The kinematic arguments are shown to lead to (1) the law of inertia for a substratum (i.e., a density distribution satisfying the cosmological principle), (2) the existence of "gravitational" mass, and (3) the inverse-square character of "gravitational force." Further problems which can be treated by the methods outlined in this paper are also indicated. The importance of the two time scales, that of t and that of τ, in answering questions about the universe (e.g., whether the universe is expanding or is nonexpanding) is pointed out.

Cosmological theory is concerned with the distribution of matter and motion in the universe as a whole. Any particular cosmological

[1] This lecture was delivered at the Symposium on Galactic and Extragalactic Structure, held in connection with the dedication of the McDonald Observatory on May 5–8, 1939.

theory attempts to indicate the relationship between the distribution of matter (the density distribution) and the distribution of motion (the velocity distribution relative to a specified observer). Or, to put it in another way, a cosmological theory may attempt to predict the main outlines of *motion* in the universe from a knowledge of the main outlines of the density distribution; or, more ambitiously, to show why a certain main outline of motion-distribution should eventually lead to some particular—say, the observed—distribution of matter.

There are, fundamentally, two different possible starting points for an investigation of this kind—two possible modes of attack. One is to suppose the investigator to be equipped with all the knowledge, empirical (that is, experimental and observational) and theoretical (that is, deductive) that has been acquired in the course of the age-long history of science up to the present day; to be in possession of all the relevant laws of nature; and to attempt to combine these various items of knowledge just as in attacking any other scientific problem, by making initial hypotheses, by applying to these the accepted laws of nature, and by determining whether the deductions agree with observation. If they do so, the initial hypotheses may be presumed to be in some measure confirmed; or, at least, the way is open to an investigation as to whether any other initial hypothesis would give an equally successful agreement with observations. The relevant laws of nature for such an investigation will include the laws of motion (of dynamics) and of gravitation, bound together by the principle of relativity. I use the latter phrase out of deference to the reader's possible predilections, although I confess that I do not know what the principle of relativity really is, as I do not, in my own work, make any appeal to, or application of, it. Still, though nowadays the principle of relativity tends to be lost in wider and more general modes of thought, "relativity," the word itself, stands for a technique, a calculus; and the calculus will most probably be utilized by the investigator who looks at it from the point of view which I have described. This calculus, or technique, includes the attribution of intrinsic properties to "mere space," the combination of the concept of space with the concept of time to yield the concept "space-time," and the linking of the intrinsic properties of this con-

cept "space-time" with the occupation of space by matter and with the motions of this matter that ensue within it. Accepting, then, all these doctrines and points of view, the investigator will proceed to attack the problem of the universe as a whole as if it were any other dynamical problem. He simply broadens the field of investigation until it includes the totality of existent matter. He synthesizes his solution out of locally acquired knowledge of nature; and he is entitled, in this mode of solution, to make use of all his sources of knowledge, however obtained.

It will be observed that in such a solution three types of knowledge are involved: intuitive, deductive, and empirical. Among items of intuitive knowledge is the awareness of the passage of time for the individual observer, about which I shall have more to say later. Among items of deductive knowledge are the consequences of the notions of number, of point, line, and plane, and, in short, the theorems of arithmetic, algebra, and geometry. Among items of empirical knowledge is the empirical fact that the number of spatial dimensions or degrees of freedom of a point is three, together with the empirical laws of dynamics and gravitation, as already mentioned. The empirical laws of dynamics and gravitation have, of course, a different status from such a brute fact as the number of spatial dimensions. Their "sanction," i.e., the authority behind them, is their verification in a large number of concrete instances, or, rather, the verification of their consequences when they have been combined with deductive elements. That is to say, they are inductive generalizations based on the supposed "principle of induction," which is so difficult to state and still more difficult to establish.

There is, of course, nothing illegitimate in this, the first, method of attacking the cosmological problem. It is the method adopted for all smaller problems. Why, then, should we not be content with it for the grandest of all problems? The answer is that, even if it can be applied completely, it leaves our curiosity largely unsatisfied, for why should an observer situated on a body, a member of the universe, have to become aware of the results of a large number of small-scale experiments before he can understand the spectacle presented to him? The spectacle itself constitutes a grand experiment in being. Should not this experiment tell its own tale and point its own moral?

Moreover, to say that a large-scale system can be understood because it exemplifies principles known to be true in small-scale systems gives us no insight into why these principles hold good at all. To take the principles as given is to admit an element of magic, of the irrational, in the working of the world, for if they have no further exaplanation, they are just magical. It is the claim of the general theory of relativity that it rests on observation and experiment —for example, on the fact that inertial mass and gravitational mass are equal. In so doing, it necessarily admits, great as are its achievements, that it is not fundamental, for it is always possible to ask deeper questions, such as "Why is gravitational mass equal to inertial mass?" and "Why should the occupation of space by matter cause what relativists call 'curvature of the space'?" We are entitled, it appears to me, to ask these and many other deeper questions, such as "Why does such and such a law of nature, supposedly irreducible to any simple law, hold good?" "What is meant by the 'rigid body,' or the 'uniform time-keeper,' of relativistic metrology?" "Why should there be an invariant measure of the separation of two events?" And so on.

It was a view held by Mach, and since assented to by many thinkers, that some, at least, of the laws of nature, such as the law of inertia, are consequences of the distribution of matter in the universe; that is to say, it is not that the motions of particles in the universe ensue in virtue of (*a*) the distribution of matter in the universe, together with (*b*) certain laws of nature; but that (*b*), the laws of nature themselves, hold good simply in virtue of (*a*), of the particles being in the presence of the rest of the universe. If there is any validity in such a view, it can be ascertained or tested only by actually applying this view, that is to say, by refraining from assuming at the outset those laws of nature whose origin is in question. This leads us to the second main approach to a solution of the cosmological problem, which is to study the universe as it actually presents itself to us, without appealing to any small-scale experiments or to knowledge derived from small-scale experiments. That is, in short, to use only such brute facts, such irreducible facts, as are of the intuitive sort or do not rest on the questionable principle of induction, and thus to appeal to no empirical "laws of nature" of a quantitative kind. The

difference between the two methods can be appreciated by a comparison with the theory of numbers or with geometry. A theorem in the theory of numbers may be discovered first empirically, and the weight of evidence for it may be such as to compel a belief in its general validity; but no mathematician accepts this as the last word on the subject. He essays to prove it, that is, to establish it deductively from axioms which, in effect, are definitions of the subjects of discourse, without appealing to any empirical evidence. It is not that he despises or rejects the empirical evidence; the empirical evidence is most welcome as a verification of the absence of logical flaws from the details of the process of inference, besides its initial role in discovering the theorem. But the theorem cannot be held to be connected with the discipline which is the theory of numbers so long as the evidence for it is wholly or partly empirical. The mathematician must, at some stage, perform an act of renunciation and proceed without reference to empirical facts. Some empirical facts, like Fermat's "last theorem" and Riemann's hypothesis, have so far defied proof; but this does not mean that we are to regard them as inaccessible to proof.

Similarly, no one now tests a geometrical theorem empirically—no one, for example, publishes memoirs of experiments designed to ascertain how exactly the three perpendiculars to the sides of a triangle from the vertices meet in a point. Yet, presumably in its pre-Greek stage, geometry passed through such a phase in which theorems were discovered or known empirically. It remained for the Greeks to invent new and more powerful combinations of processes of inference which were capable of establishing the empirically known theorems deductively, from axioms which (in their modern usage) are concealed definitions.

There are, then, these two separate ways of attacking the structure of the universe: the one to make use of every available piece of empirical knowledge known to be valid on the small scale; the other to begin with the situation actually presented to us by the totality of things without supposing ourselves to know anything, to start with, of the facts used in the first method. In the second method we attempt a complete reconstruction of physics from the bottom upward, on an axiomatic basis. It might be supposed that the sec-

ond method, starting with a poorer armory of resources, could never get as far as the first method. Actually, it proves to be more powerful, for, as it proceeds, it frees itself from the presuppositions tacitly made in the first method.

I will now say something about the results of the two methods. In principle, the natural way of applying the first method would be to assume some distribution of density in the universe, apply the laws of dynamics and gravitation and calculate the resulting motion. It is natural, in the first instance, to treat each extragalactic nebula as a particle and to mean by "density at any locality" the average number of such particles per unit volume; the system is then compared to a dust cloud, in which the separate particle-members are in free motions under another's gravitational influence. The motion will then depend on the initially assumed density distribution. The method followed by general relativity is, however, indirect: it first assumes a metric (described picturesquely as "some kind of expanding space"), determines the motions implied, and then by means of dynamical equations (field equations) calculates what uniform density (spatially constant but a function of the time) is implied by this set of equations.

It is a remarkable fact, pointed out in 1934 by McCrea and the writer, that the set of equations thus obtained connecting density and motion, obtained from the field equations of general relativity, are identical with those deduced from a crude nonrelativistic Newtonian theory in which the main role is played by the Newtonian equation of motion of a particle under the gravitational influence of the matter inclosed within the sphere through the particle centered at the observer and a hydrodynamical equation of continuity. It then appears that there are three types of motion—motion with a velocity equal to, exceeding, or less than the "velocity of escape" from the sphere of matter just mentioned—and that these three types of motion correspond exactly to the three possible types of "expanding space," namely, of zero, negative, or positive curvature. The interpretation of the equations in terms of observations is different in the Newtonian and relativistic schemes, for they employ space in different senses. Nevertheless, the general dynamical content of the two schemes is essentially the same. And the astounding

fact emerges that neither of these schemes, which both follow the first method of our general discussion, contains any criterion by which one can decide which to expect in nature, out of the three supposed possibilities. General relativity is powerless to tell us whether the space in terms of which it analyzes the universe is expected to be of zero, negative, or possible curvature, whether the motions to be expected are of parabolic, hyperbolic, or elliptic type; whether the universe is expected to be expanding forever, to be oscillating between finite nonzero radii, or to be expanding and then contracting to zero. No theoretical criterion has been found capable of distinguishing between what are, to general relativity, fundamentally different possibilities.

Science has long been acquainted with problems of the type that cannot be solved. Such have in the past been, for example, the problem of perpetual motion, the problem of squaring the circle, of trisecting an angle. It has usually been proved in such cases not that man's intellectual resources were incompletely applied but that some erroneous tacit assumption has been made in the formulation of the problem, or, again, that the problem was a no-problem, that it did not exist. After men had striven for years, for example, to prove Euclid's parallel postulate, it occurred to the founders of non-Euclidean geometry to do without it altogether—in short, to deny it. Is it not probable that this continued failure of general relativity to solve a fundamental problem proposed by itself means that there is something wrong with the formulation of the problem, or with its embodiment in analysis, or with the assumptions tacitly made in stating the problem? I invite you to consider the most attractive alternative: that all these differing possibilities are different descriptions of the same entity—descriptions which originate from the tacit adoption of different scales of time; that there is one universe but many different possible descriptions of it; that there is actually no diversity of possibility of the thing in itself, but only a diversity of possibility of description. This will lead us directly to the second mode of attack on the cosmological problem.

One of the first questions that can be asked about a system of moving particles is whether it is in a steady state or not. If it is in a steady state, its statistical description at any one instant is the same

as its statistical description at any other instant; the variable t cannot enter into the description—one moment is as good as another. But if it is not in a steady state, the state at one instant distinguishes itself from that at another. It should, therefore, be possible to indicate to an intelligent person which instant is being the subject of discussion or examination. When the same problem occurs on a less grand scale, when the person is interested in a system which is some small part of the universe, the same problem will occur; but, then, the person will have some experience of other small systems, into which experience he can fit his experience of the given small system. In other words, once his scale of time is fixed by some one of his individually experienced systems, he can correlate his experience of other systems with that one. That one system constitutes his clock, and he can use it to record the evolution of other systems open to his observation. But if his field of observation is the whole universe, and, following our second method, he renounces appeals to smaller systems, that whole universe must itself constitute his clock. And as one aspect of the spectacle succeeds another and so distinguishes itself from another, so the observer must be able to differentiate his description of it and refer them to different instants in his clock scale. To have a meaning, these must be communicable to any other intelligence, and, therefore, to describe them he must be able to refer to some natural time-zero inherent in that description of the universe which he is trying to communicate. The universe, if not in a steady state, must thus contain within its own description some identifiable time-zero. On the other hand, if the universe is in a steady state, any epoch can be taken as time-zero; past and future are illimitable; and the "earliest possible" moment recedes to minus infinity on this time scale.

I confess that this situation has always created for me a difficulty in the well-known paper which Hubble and Tolman published in the *Astrophysical Journal* in 1935.[2] In that paper a certain function, $g(t)$, was expanded in powers of t in such a way that the epoch of observation was taken as $t = 0$. But this places all epochs of observation on the same footing; it bars the development of the system of the universe *in time*, for it contains no specification of what mo-

[2] 82, 302.

ment is the moment $t = 0$. In spite of the red shift, then, the procedure is tantamount to discussing only a stationary universe, in the sense that the state contemplated at any one epoch of observation is the same as that contemplated at any other epoch. To remove this restriction it is necessary to incorporate into the analysis some parameter corresponding to a zero of time, i.e., to some recognizable event in the evolutionary history of the subject; and then to relate any present epoch with the introduced zero, by reckoning it from that zero. In the investigations contained in my book *World Structure*, which were parallel to those of Hubble and Tolman, I always left in a time variable t, which could measure the epoch of observation reckoned from some well-defined zero. It is the absence of a t in Hubble and Tolman's final formulae which makes me suspect that their analysis describes only a stationary system. Not that it is any the less apropos for that, as I hope to show.

This forcing on our notice of a zero of time, this recognition that the universe itself must ultimately be the clock to be used in the second main method of approach to the cosmological problem, leads us to consider in more detail the problem of time as it occurs in science.

In relativity the notion of uniform time is adopted without discussion. It is said that a material clock is constituted by any "regularly periodic phenomenon," though how one is to recognize when a phenomenon is strictly periodic and when not is not stated. Suppose I have two clocks, side by side. One contains a mechanism which makes it gain on the other by an amount which, starting from the epoch when the clocks agreed, varies as the square of the epoch read by the other. Which measures uniform time? No one can say. No one can catch the fleeting minute and set it side by side with another supposed minute. Again, we can, if we like, say that uniform time is that which makes the laws of dynamics hold good. But this is equivalent to graduating one clock—say that given by the rotating earth—in a particular way and calling that way "uniform." Obviously, it is no more entitled to the word "uniform" than many other possible ways of graduating the clock. We have to recognize that the word "uniform" in the context of "uniform time" is one of those emotion-soliciting words which have no place in science, and

we must attempt to proceed without it. All modes of graduating a "clock" are then on the same footing, to begin with.

The word "uniform" in the phrase "uniform time" has the same dangerous plausibility as the word "rigid" in the phrase "rigid body." Most people feel that they have an intuitive sense of what is meant by a "rigid length scale," just as they believe themselves to have an intuitive sense of what is meant by "uniform time." But no one can say offhand what is meant by either phrase. I will attempt to illustrate by an example. It is a commonplace that *motion* is relative. It is perhaps less of a commonplace that *position* in the universe is relative. It is often asked whether the distant nebulae are "really" receding from us or not. I shall show later that this question has no meaning. For the moment, let us see what the propounder of the question usually means. He means to inquire: If we could put up a fixed "post," say a "bench mark" in space, at a fixed distance from ourselves, in the vicinity of a nebular nucleus, would the nucleus be passing the "bench mark" or not? To make sure that the bench mark is at a fixed distance from ourselves, we must be continually measuring its distance, ideally with "rigid" meter scales. Suppose these are put end to end, so as to join ourselves rigidly to the bench mark. Now suppose a uniform extensible rope joining ourselves to the nebular nucleus, with the marks on it coinciding, at a certain instant t, with the meter markings of the rigid scale. Then the problem of whether the nebular nucleus is "really" moving or not is the same as the problem of whether the first meter marking on the rope is or is not passing the first meter marking on the "rigid" scale. If the nebular nucleus is supposed to be moving at the rate suggested by Hubble's determination of the red shift—say 500 km/sec/10^6 parsecs—then the first meter marking on the rope will be moving past the first meter marking on the "rigid" scale at the rate of

$$\frac{500 \times 10^3 \times 3.16 \times 10^7}{10^6 \times 3.08 \times 10^{16}} \text{ meters per year},$$

or by one part in 2×10^9 per year. It is sometimes supposed that the nebulae are receding (if they are receding at all) very fast. It appears from this illustration that, considering their distances, they are re-

ceding excessively slowly. In fact, we could rarely distinguish between the supposed fixed length of the meter scale and the supposed expanding first segment of the expanding rope. But if so, how can we be sure that the supposed "rigid" meter is not "naturally" expanding with time? And again, if so, we could join the nebular nucleus to our own galaxy with a rigid rod without the latter's snapping. But that would be the same thing as the nebula's not being in receding motion at all. It is just when we come to very small rates of motion like that of the nebulae that it is of supreme importance to be clear what we mean by a "rigid length scale." It is as imperative to define the latter as to define "uniform time." But both are indefinable, conventional expressions. The only thing left to do is to link the two.

We can do this by means of the to-and-fro transmission of light. Having chosen a mode of graduating our clock, we shall say that a given particle is distant a "constant distance" from ourselves by this clock, when the interval between the transmission of a light-signal and the reception of the reflection signal is constant, as recorded on this clock. It is to be observed that we assert nothing about the velocity of light. The velocity of light is, in fact, quite indefinable until we have specified a clock and given a definition of a distance. Thus, we are not assuming that the "velocity of light is constant." Nor do we assert that the forward and backward velocities of light are equal —another old bogey in the supposed assumptions of relativity, for we have, as yet, no means of specifying the epoch of arrival of the signal at the distant particle—we have so far constructed no "clock at a distance." We have, however, linked our distance measure with our clock graduation.

We put this in symbols. A "clock" is, for us, just some arbitrary correlation of the events of our experience with the real numbers. Such a correlation is possible because the events of an observer's experience form a well-ordered sequence—ordered by the "before-and-after" relation. As Shakespeare says, "We see which way the stream of time doth run."[3] (I will remark at this stage that there is no actual need to restrict the correlation to the *real* numbers. Any *linear* continuum of numbers, real or complex, will serve to order the well-ordered sequence of events in the observer's consciousness, and

[3] *Henry IV*, Part II, Act IV, scene 1.

it actually occurs in certain cases of regraduation of clocks that it is convenient to adopt a complex time scale.) Our clock, then, is at ourselves and records only the epochs of events at ourselves. If, therefore, t_1 is the epoch of the transmission of a light-signal to a distant particle, and t_3 the time of return, we put $\frac{1}{2}c\,(t_3 - t_1)$, where c is an arbitrary constant, as the distance of the distant particle by this clock, i.e., by this mode of clock graduation. We do not, of course, know or care whether this assessment of "distance" agrees with any other measure of distance that may be made by some other method; it suffices that we have a measure of the separation of the distant particle from ourselves at a certain definite instant. What instant? The instant of reflection. How are we to give a measure to this? Out of our two independent observations t_1 and t_3 we can construct two equivalent data in an infinite variety of ways. One we have already constructed, namely, $\frac{1}{2}c(t_3 - t_1)$. Let us adopt for the other $\frac{1}{2}(t_3 + t_1)$ and call this the epoch, by the same clock, of the event of reflection. We put $r_2 = \frac{1}{2}c\,(t_3 - t_1)$, $t_2 = \frac{1}{2}(t_3 + t_1)$.

It is clear that the measure of epoch-at-a-distance, t_2, equal to $\frac{1}{2}\,(t_3 + t_1)$, is largely arbitrary. We might equally well have chosen $(t_3t_1)^{1/2}$ or $[\frac{1}{2}\,(t_3^2 + t_1^2)]^{1/2}$, etc. This is a familiar topic in relativity —the conventionality of assignment of an epoch-at-a-distance. But this shows up at the same time, strikingly, that the assignment of distance-at-an-epoch is equally conventional. Instead of taking $\frac{1}{2}c\,(t_3 - t_1)$, we could take $\frac{1}{2}c\,t_0 \log\,(t_3/t_1)$ or $\frac{1}{2}c\,(t_3^2 - t_1^2)^{1/2}$, etc. This is a *less* familiar topic in relativity. General relativity is accustomed to regard all co-ordinates and combinations of co-ordinates as on the same footing. It is then prepared to take any function of t_3 and t_1 as a co-ordinate defining "epoch-at-a-distance." In particular, we may replace t_3 and t_1 themselves by T_3 and T_1, where $T_3 = \chi(t_3)$, $T_1 = \chi(t_1)$, and the T's denote clock-readings regraduated. How then, if all co-ordinates are on the same footing, does it succeed in picking out one mode of clock graduation as unique? It does so by just the same process as it uses when it claims to pick out one mode of distance measure as unique—in both cases by refusing to define, owing to the impossibility of defining, a uniform "clock" or a "rigid" scale. The result is that it cannot be sure that one of the co-ordinates it employs is identifiable with the clock time of physics, or that an-

other is identifiable with the supposed rigid "length measure" of physics.

I do not propose to contest ground, inch by inch, with general relativity. (It makes so many disconnected assumptions that it can usually extricate itself from the inadequacy of any one of these by relying on some other.) Instead, I hope I have succeeded in showing how fundamental in a discussion of the cosmological problem is the fixation of a distant particle's state of rest or motion relative to ourselves, and that the description—i.e., "as a state of relative motion" or "as a state of relative rest"—depends partly on the way our clocks are graduated and partly on conventions.

Closely connected with the problems of setting up a clock at ourselves and of choosing our conventions for epoch-at-a-distance and distance-at-an-epoch is the problem of setting up a clock B at a distance which shall be a copy of the clock A set up at ourselves. (We shall see later why this has an important bearing on the cosmological question.) The clock at a distance B is supposed to be represented by the again arbitrary graduation of the temporal experience of the observer at B. This observer is identifiable with some particle which may be "moving" in any manner, its motion being assessed by our clock A, by specifying the relation of its distance-at-an-epoch to the epoch concerned. How can A dictate instructions to B, enabling B to graduate his clock (or, if already graduated, to regraduate it) so that B's clock may be said to keep the same time as A's? Clearly, on the possibility of doing this rests the whole significance of speaking of "time" at one place in relation to "time" at another.

If A's and B's experiences are completely uncorrelated, no meaning can be attached to asking whether they both "keep the same time," for their experiences would be independent. Their experiences may be correlated, in principle, merely by allowing them to view one another—in addition, that is, to allowing them to assess one another's motion by to-and-fro light-signals. Now, if there is any asymmetry in the relation holding between A and B, we could not deny the possibility of an asymmetry in their description of one another, each using his own clock; and we should not know how much of this asymmetry in description was due to an asymmetry in their respective methods of time-keeping. If, on the other hand, the relation between

A and B is a symmetrical one, it must be possible for them to graduate their clocks so that the symmetry extends to their descriptions of one another when made by each using his own clock and viewing the other's. When such graduation is achieved, they may be said to be "keeping the same time." More precisely, it can be shown that A, having arbitrarily graduated his own clock, can dictate instructions to B (derived by A from his observations of B) so that B can regraduate his clock to keep the same time as A's. The clocks are then said to be congruent, the observer's equivalent. It can be shown further that, if at any time during their relative motion, A and B coincided momentarily in position, then B's regraduation of his clock can be uniquely determined. (This is connected with the circumstance that, when the two observers coincide, their clocks must read the same epoch. But it is of a much more subtle character than this circumstance alone would appear to indicate.)[4] If A and B never coincide, there remains an arbitrariness in B's regraduation.

Now that A and B keep the same time, we may inquire under what circumstances a third observer, C, keeps the same time as both A and B. It is convenient to confine attention first to observer C collinear with A and B. Then, to state the problem precisely, we may ask whether, when A has dictated instructions to B and to C, so as to put them in possession of clocks congruent to A's, B and C will find themselves in possession of clocks congruent to one another. The answer is, in general, that they will not. C's clock, in these circumstances, will be congruent to B's only if there is a certain restriction placed on C's motion—his motion, that is, relative to A or B. It can be shown further that when C satisfies this condition, if a fourth collinear observer D is considered, and it is asked that he shall have a clock congruent to A's, B's, and C's, then the motion of D, and a fortiori of all other collinear observers E, F, ... , is completely determinate. Such a collinearity of observers possessing congruent clocks is called a "linear equivalence." Its study is fundamental in all questions of time-keeping and in the general problem of the motion structure of the universe.

A property of a linear equivalence is that, if two members ever

[4] See a paper by M. Crum communicated to the *Quarterly Journal of Mathematics* (Oxford). (**10**, 155, 1939.)

coincide in position at some epoch, then all members of the same equivalence coincide in the same position at the same epoch. But there need be no epoch of coincidence. Examples of linear equivalences are: (1) a set of stationary particles, (2) a set of particles in uniform relative motion which all coincided at one epoch; (3) all kinds of accelerated motions of particles starting from coincidence, among which is the equivalence of particles in so-called "uniform" acceleration considered by Leigh Page; (4) all kinds of accelerated motions of particles which never coincide.[5]

But, however much these types of motion may appear to differ, they are essentially one and the same entity, in that by appropriately regraduating any one observer's clock in any one linear equivalence according to some law $T = \chi(t)$, and then regraduating all other observers' clocks according to the same law, this linear equivalence may be reduced to any other selected linear equivalence. That is to say, the descriptions of the motions of the members of the first equivalence as seen from any one member and determined by means of his clock can be brought into coincidence with the description by any other member of any other equivalence, by suitable clock regraduation. It may occur that in certain cases the regraduation is complex, but this does not affect the principle.

By suitable means the linear equivalences may now be generalized to give three-dimensional equivalences, which fill three-dimensional space. We then have, as examples, the relatively stationary equivalence, consisting of a three-dimensional set of particles at relative rest and a set of particles in uniform relative motion expanding outward from any one of them, considered as origin. We can also have any variety of accelerated motion. In accordance with the general principle, these can all be reduced to one another by clock regraduation. For example, if we regraduate the clocks of the uniform motion equivalence, reading time t, by a logarithmic regraduation to time τ, where $\tau = \log t$, or, better, $\tau = t_0 \log (t/t_0) + t_0$, where t_0 is an arbitrary constant, then the uniform motion equivalence turns into the one at relative rest.

There is, in accordance with expectations mentioned earlier and

[5] It has been shown by Whitrow that the general linear equivalence is characterized by a generating function ψ, such that $t_2' = \theta_{AB}(t_1) = \psi a_{AB} \psi^{-1}(t_1)$, where t_2' is B's time of reception of the signal leaving A at A's time t_1, and a_{AB} is a number.

based on general principles, a natural time-zero for the uniform motion equivalence, which may be taken to be the epoch of coincidence of all the members. In accordance with the same general principle, the relatively stationary equivalence, being in a steady state, can have no zero of time. We notice at once, in confirmation, that $t = 0$ corresponds to $\tau = -\infty$. By adjustment of the arbitrary constant t_0, τ can be made to take any value for any given value of t and, in particular, to agree with t when t takes the value t_0.

An equivalence is simply a distribution of motions, the particles executing which have a common measure of time and carry congruent clocks. For a world-wide system of congruent time-keeping it appears necessary that the clocks shall form members of an equivalence. The fundamental problem therefore propounds itself: Are the extragalactic nebulae, considered as particles, members of an equivalence?

If they are, several other questions answer themselves at once. For example, though an equivalence may be described in different ways, according to its mode of clock graduation, there is essentially only one equivalence. Hence, if the universe is of the motion type of an equivalence, there is only one type of universe possible. This is at once an immense improvement on the answer given by the older, relativistic (first-method) theories, which, as we saw, led to different possible types of universes, among which they could not discriminate. We have, instead, different possible descriptions of one and the same underlying kinematic entity. The allegedly different motion types of the older theories must be due either to their failure to discriminate between the effects of different types of clock graduation in altering the description of a system or in their admission of possibilities, such as that a light-signal may be received back at two distinct epochs, which we have not admitted. Such possibilities, which lead to such profound difficulties on further analysis, may be discarded; they have led to such fantastic results as that a particle may affect the motion of another particle the more the farther away it is —results repugnant to common sense and easily seen to derive themselves as mathematical rather than physical possibilities. It is highly satisfactory that, when we approach the problem of the structure of the universe by the second method, we are led to only one eventual

possibility, different as are the appearances which this possibility may take. It has been sometimes claimed as an accusation of lack of generality in the present treatment that it does not provide these tiresome choices of $k = 0, +1, -1$, which relativists so often dwell on. But we have been led to a single possibility inevitably and not by ignoring choices of possibilities presented to us. After all, there is only one universe; and a satisfactory analysis of the possibilities of universe structures should lead to only one type of such structure.

Again, the objection has been raised to the present solution that it is "special," in the sense in which the restricted theory of relativity is special compared with the so-called "general" theory; that it contemplates only uniform motion, and handles it only by Lorentz formulae. But this does not impugn the generality of the system discussed, for, according to the general theorem, we can reduce any motion type to uniform relative motion, any equivalence to the equivalence in uniform relative motion. We can, therefore, handle the most general type of motion, involving accelerated relative motion of any character, by suitable regraduation of clocks; and the formulae of transformation among members of any equivalence, easily obtained in their general form, pass, by suitable regraduation of clocks, into Lorentz formulae. Actually the present analysis of the cosmological problem is perfectly general; it is simply that I choose a particular idiom in which to describe it, instead of using a characterless general language.

If the universe is representable as an equivalence, we see at once that the question often asked, "Is the universe expanding or not?" is a no-question. For an equivalence can equally well be regarded as relatively stationary or as in any kind of relative motion; in the former case there is no epoch of coincidence of all the members, in the latter case there may be. Simply by regraduating clocks, the same equivalence may be represented as of the one character or the other. Thus, whether the nebulae all coincided at one epoch or not is another no-question. Certain modes of clock graduation lead to special features in the resulting description. For example, the only mode of description which contains no parameter t_0 or its equivalent, in itself, is that corresponding to uniform relative motion. The only one containing no natural zero of time is the relatively stationary one; and

in this the epoch assigned by any observer to an event is the same whatever the observer. (It should be particularly noted that, when we are considering equivalences, the relatively stationary equivalence is not a particular case of the uniform relative-motion equivalence.)

It thus becomes necessary to ascertain observationally whether the system of the extragalactic nebulae constitutes an equivalence—whether, that is, there is a world-wide possibility of the construction of congruent clocks. This is a question logically quite anterior to the question which has so preoccupied the minds and time of astronomers, namely, whether the universe is homogeneous or not. For it should be carefully noticed that in the specification of an equivalence no density distribution is concerned; density of members is entirely irrelevant. The specifications of an equivalence involve simply the motion of each individual member relative to some given member—the relation, that is, of distance-at-an-epoch to epoch for that member. We want to know, for the extragalactic nebulae, whether they belong to the equivalence defined by any three of them.

If we could dispatch to and receive light-signals from them, these would be, in principle, a simple but slow matter. Actually we can only *receive* signals. And the building-up of these signals into the conventional numbers called "distances" and "velocities" is a very difficult business if we are to relate them to fundamental measures made with a given clock. The methods of determining nebular distances involve as an intermediary the construction of a good deal of geometry and a good deal of physics; and the further discussion of motion involves the differentiation of these distances, when ascertained, with time. The observation of the red shift by itself tells us nothing save that, to a first approximation, atoms keep kinematic time—the time of a clock graduated to represent the equivalence as being in uniform relative motion; and any equivalence obeys to a first approximation a linear relation between distance and velocity. One of the greatest of theoretical needs is the determination of what a Mount Wilson or Harvard distance means on the fundamental time scale it must be tacitly based on.

And one of the greatest of observational needs is the detection of absolute change in the geometrical appearances of extragalactic objects with time. It seems to me that here is one of the principal tasks

of large telescopes. The most distant visible objects, increasingly dimmed by recession, but more and more numerous, will have enormous red shifts. But they will be at a finite distance. They will therefore give the maximum chance of yielding changes of angular diameter with time, for example. I have not the slightest doubt that no such changes will ever be detected, even if observationally detectable in principle. For, if the universe can be assimilated to a three-dimensional equivalence, it can be regarded as a stationary universe, in which an angular diameter remains constant. The interpretation of this on the kinematic scale, i.e., on the scale corresponding to uniform relative motion, is that the linear dimensions of the nebulae, on that scale, are expanding in proportion to their distances. The same consideration applies to a so-called "rigid" length scale. For, if one is supposed erected at a nebular nucleus, perpendicular to the line of sight from ourselves, it must subtend a constant angle in a stationary universe on the τ-scale, and therefore also on the t-scale; hence, since its distance away is increasing on the t-scale, its measure of length on the t-scale is increasing. This again shows that, if a nebular nucleus is supposed joined to ourselves by a chain of "rigid" rods, it will not part company from the end of the chain on the τ-scale, and so not on the t-scale; and hence the rods must be expanding uniformly on the t-scale.

The fact that the red shift is proportional to distance is, as remarked, evidence that the nebulae compose an equivalence and that atoms keep approximately kinematic time as regards the frequencies they emit. It does not tell us whether the orbital frequencies of the atoms are constant in any particular scale. That question takes us to dynamics, and with it to the distribution of density in the universe—an important question, but one coming second in logical order. Before, however, we can discuss atomic dynamics, we must have a dynamics capable of general applications. Such a dynamics must be based on the motion of a "free" particle, for a force can be measured only in terms of departure from "free" motion. Examples of "free" particles are afforded by the nebular nuclei themselves. So far we have considered these "motions" or states of rest kinematically, i.e., as something given. We can now go on to ask whether an equivalence represents a natural motion, i.e., whether the members of an equivalence can of themselves, without an "external" agency,

pursue the trajectories provided for them in the specifications of an equivalence. To solve this problem it is first necessary to consider the motion of a free particle in general in the presence of an equivalence. And we choose, for the scale of time, i.e., the mode of clock graduation in which to conduct our analysis, the t-scale, the one in which the members of the equivalence are in uniform relative motion.

Can we assert anything about the motion of such a free particle? Clearly, all we can do is to pass from one member of the equivalence, considered as point of observation, to another—from one "fundamental" particle to another. We can transform the equation of motion from one origin to another. The new form which this equation of motion takes will depend on the distribution of the totality of particles constituting the equivalence. For example, the equation of motion of a free particle in a dense part of the equivalence might be of a different form from that of a free particle in a less dense part of the equivalence. For we must remember that, when we speak of a "free" particle, we mean a particle in the presence of the totality of matter in the universe, or, as we should normally say, a particle exposed to the "gravitational pull" of the rest of the universe. We can assert nothing whatever about the motion of a particle "ignoring" this effect of the rest of the universe, for the rest of the universe is necessarily present—without it there would be no means even of describing or locating the "test" particle. Thus, the dynamics of a "free" test-particle must depend on the density distribution in the equivalence. Until the latter is specified, the "dynamics of a free particle" has no meaning. This density distribution of the equivalence is something quite different from the motion pattern which defines the nature of the equivalence in the chosen scale of time. For the same motion pattern we can have kinematically any kind of density distribution. The specifications are of a quite different character, and the specification of the motion pattern comes logically before that of the density distribution. The question so often discussed, "Is the universe homogeneous?" is logically posterior to the question which ought to be asked first, "Is the universe an equivalence?"

Let us then suppose we have specified our equivalence and specified a density distribution in it. Two questions now arise: First, can

we determine from this the equation of motion of a free test-particle in the presence of this density distribution? Second, do the given particles of the equivalence obey this law of motion; i.e., is the equivalence, when kinematically specified, also a dynamical system itself?

These questions have not yet been answered save in one important case, namely, that in which the equivalence is a substratum. A substratum is an equivalence whose density distribution satisfies the cosmological principle; that is, if A and B are any two members of the equivalence, then the statistical description of the whole equivalence, from A, in its temporal unfolding, coincides with that from B. This we write $A \equiv B$, a statement to be carefully distinguished from $A \equiv B$, which means simply that A and B have congruent clocks. When $A \equiv B$, for any two members A and B, the system constitutes an equivalence; when, in addition, $A \equiv B$, it constitutes a substratum. The ideas contained in the assertion $A \equiv B$ are a generalization of the ideas contained in the notion of homogeneity. Whether a substratum appears homogeneous or not depends on the scale of time adopted; for example, on the τ-scale, or the scale in which the equivalence appears as at relative rest, a substratum is a homogeneous distribution of fundamental particles; but on the t-scale the density distribution will appear to increase outward from any arbitrary observer, although no observer is in a favorable position—a remark which, I think, removes some of the difficulties stated in the 1936 Rhodes lectures by Hubble.

It is clearly necessary to solve this, the fundamental problem of dynamics—the determination of the motion of a free particle—for an equivalence satisfying the cosmological principle before attempting it for any other. This must afford a norm of knowledge, a standard of expectation. When we attack this problem analytically, when we attempt to find the equation of motion of a free particle in the presence of a substratum, we find two highly significant results.

The first is that the answer appears, at first sight, to be indeterminate. The second is that, in spite of the indeterminacy, the kinematically prescribed motions of the particle members of the equivalence satisfy the dynamical equation of motion.

Choose the scale of time in which the equivalence $(A \equiv B)$ is in uniform relative motion, that is, the t-scale. Then the motion pat-

tern is $\boldsymbol{V} = \boldsymbol{P}/t$, where $\boldsymbol{V}$ is the velocity of the fundamental particle at $\boldsymbol{P}$ at time t. The corresponding substratum has a density distribution

$$n\,dx\,dy\,dz = \frac{Bt\,dx\,dy\,dz}{c^3\left(t^2 - \frac{\boldsymbol{P}^2}{c^2}\right)^2}, \tag{1}$$

provided that, in addition to the cosmological principle ($A \equiv B$), a certain other principle, called the "principle of symmetry," is satisfied.

The equation of motion of the particle passing through $\boldsymbol{P}$ at time t with velocity $\boldsymbol{V}$ is

$$\frac{d\boldsymbol{V}}{dt} = (\boldsymbol{P} - \boldsymbol{V}t)\frac{Y}{X}G(\xi), \tag{2}$$

where $Y = 1 - \boldsymbol{V}^2/c^2$, $X = t^2 - \boldsymbol{P}^2/c^2$, $Z = t - \boldsymbol{P}\cdot\boldsymbol{V}/c^2$, and $\xi = Z^2/XY$. The function $G(\xi)$ is not determined by the two conditions $A \equiv B$ and $A \equiv B$.[6] But whatever the form of $G(\xi)$, the fundamental particles moving according to $\boldsymbol{V} = \boldsymbol{P}/t$ satisfy the equation of motion (2).

This is a result of fundamental importance, whose interpretation has, I think, not so far been given. It means, I think, that the kinematical system called an "equivalence" can only be also a dynamical system when it is also a substratum satisfying $A \equiv B$. For suppose that we now alter the density distribution so that it no longer obeys the cosmological principle, but that we leave the motion pattern unchanged. Then we shall certainly alter the equation of motion of a free particle. This alteration will not be merely a change in the function $G(\xi)$, as we shall see in a moment. It must, therefore, be a more drastic change, and the law $\boldsymbol{V} = \boldsymbol{P}/t$ will no longer satisfy (2). Hence, the equivalence, as specified kinematically, will not conserve itself dynamically and must alter in its state of motion, not necessarily even remaining an equivalence. The substratum, or, as we may call it loosely, the "homogeneous equivalence," is therefore a fundamentally important type of equivalence in that it alone conserves its character of being an equivalence.

[6] The reader who is unfamiliar with the mathematical background may be informed that $A \equiv B$ determines the transformation from one fundamental particle to another, while $A \equiv B$ determines the conservation of form of the equation of motion and the density distribution.

There are, of course, many difficult, unsolved problems to be investigated in this region. We should like to know, for example, the conditions of stability of an equivalence arranged as a substratum; this would require the knowledge of the equation of motion of a free test-particle in the presence of an equivalence which was not a substratum—again an unsolved problem. But one remark may be made before we pass on. It is commonly supposed that there is such a thing as a "dynamics," that is, a law of nature giving the equation of motion of a free particle in a form independent of epoch and location—a law of nature satisfying the principle of the uniformity of nature. My point is that those who believe this are tacitly using the cosmological principle. If the universe is not homogeneous (in our extended usage of the word), then the form of the dynamics of a free particle will depend on locality, and there will be no world-wide dynamics possible. There is not sufficient observational evidence to say whether a world-wide dynamics is necessary or not; but it is inconsistent to appeal to, or to assume, the principle of the uniformity of nature and yet to deny the cosmological principle as soon as it has been shown that the forms of the laws of nature depend on the distribution of matter in the universe and that these forms can be the same everywhere only when the distribution of matter is describable in the same terms from every fundamental particle as observing-point.

The next thing to notice is the apparent indeterminacy contained in the function G. It may be remarked that the function G in the form $G(\xi)$ has been reduced from the more general form $G(X, \xi)$, from the consideration that the substratum, as specified by (1), contains no constant of the dimension of a time in its description; and as X could occur only in the form X/t_0^2, where t_0 was a constant, it cannot occur at all.[7] Thus $G \equiv G(\xi)$.

Now, why is G thus indeterminate? The substratum is indeterminate only to a constant B, and this is equivalent only to the arbitrariness in reckoning what is a fundamental particle: it has no dynamical significance. Hence, the equation of motion must be completely determinate—a free particle must move in an absolutely foreseeable way.

[7] On the contrary, when we use dynamical time τ, the density distribution (1) becomes replaced by $B/c^3t_0^3$; and a constant t_0 appears and may reappear (as it does) in the equation of motion of a free particle.

The contrary has been urged by H. P. Robertson in a passage which expresses the view that the universe is, at bottom, magical and irrational. Thus he remarks: "The cosmological principle reduces the problem of motion to that of determining, *by the imposition of physical law* [italics mine], a single scalar function." The view that we have to impose a physical law, i.e., something extraneous to the considerations before us, is entirely foreign to the whole set of ideas here being developed. Robertson, in fact, misconstrued the force of the kinematic argument.

The reason that G is indeterminate from the considerations so far advanced is not that G requires some extraneous "physical" law to fix it but that the system defined by specifying that it shall satisfy the cosmological principle is still itself indeterminate. A substratum, i.e., a homogeneous equivalence, is a system of hydrodynamical character; but far more general systems will have an equation of motion of a free particle of precisely the same form, simply because they satisfy the cosmological principle $A \equiv B$. The indeterminacy is $G(\xi)$ corresponds to the possibility of defining vastly more of these general systems. Since there is essentially but a single hydrodynamical system, more general systems cannot be of hydrodynamical character but must be of a statistical character. And as soon as we investigate the possible diversity of such statistical systems, we find that there is just the degree of freedom corresponding to the indeterminacy of G; namely, a single function $\psi(\xi)$, itself arbitrary, describes each statistical system. Its density-velocity distribution is found to be, in fact, of the form

$$n\,dx\,dy\,dz\,du\,dv\,dw = \frac{\psi(\xi)\,dx\,dy\,dz\,du\,dv\,dw}{c^6 X^{3/2} Y^{5/2}} . \tag{3}$$

If this exhausts the possibility of obtaining systems satisfying the cosmological principle, then we should expect that there should be a relation between $\psi(\xi)$ and $G(\xi)$. And we find, in fact, from arguments equivalent to the mere counting of particles, that the relation is just

$$G(\xi) = -1 - \frac{C}{(\xi - 1)^{3/2}\psi(\xi)} , \tag{4}$$

where C is again arbitrary. This relation is quite unlike any other relation in the whole range of mathematical physics, for without any "imposition of physical laws" it derives by purely a priori methods a relation between density and motion—in other words, a relation of purely gravitational character is derived without recourse to experiment.

We have now not only arrived at the reason why $G(\xi)$ is, at first sight, indeterminate; we have obtained an actual expression for $G(\xi)$ for the general case of a statistical or nonhydronomical distribution. By limiting processes, of which the details have been given elsewhere,[8] it can be shown that the first term, -1, in (4) represents the contribution to $G(\xi)$ of the hydrodynamical system by itself. This, therefore, determines $G(\xi)$ for a pure substratum as

$$G(\xi) \equiv -1 ,$$

which is equivalent to the purely logical deduction of the first law of motion. It is the fundamental theorem of dynamics. For, on transformation from the scale of t to the scale of τ, the equation

$$\frac{d\boldsymbol{V}}{dt} = -(\boldsymbol{P} - \boldsymbol{V}t)\frac{Y}{X}$$

becomes an equation stating that $\boldsymbol{v}$, the τ-velocity of a free particle in a certain public hyperbolic space, remains constant. It remains constant in spite of the filled-up-ness of the substratum or universe with matter, and allows fully for the gravitational effect of this matter on the particle. This shows that in a dynamics based on the assumption of Newton's first law of motion, the time variable must be what we have called τ. Our procedure is something quite different from the procedure followed in general relativity. There, the metric depends on the filled-up-ness of the space with matter, and the paths are *assumed* to be geodesies; in our procedure, the space chosen is independent of the density of the matter, and, in the t-measures, the free paths are not geodesies.[9]

As soon as the motion of a free particle is known, a dynamics can be constructed. This has been done in papers published elsewhere.

[8] *Quart. J. Math.* (Oxford), **8**, 22, 1937.

[9] In t-measure each observer has his own private space.

The principal result of interest is that *mass*, inertial mass, is represented in the t-scale by $m\xi^{1/2}$; energy, by $mc^2\xi^{1/2}$, where

$$m\xi^{1/2} = \frac{m\left(t - \boldsymbol{P}\cdot\frac{\boldsymbol{V}}{c^2}\right)}{\left(1 - \frac{\boldsymbol{V}^2}{c^2}\right)^2\left(t^2 - \frac{\boldsymbol{P}^2}{c^2}\right)^{1/2}}.$$

When $\boldsymbol{P} = 0$, i.e., for an observer momentarily coinciding with the particle, this reduces to $m/(1 - \boldsymbol{V}^2/c^2)^{1/2}$, which is Einstein's expression, save that here $\boldsymbol{V}$ denotes specifically *not* the velocity relative to any standard of rest, as in the special theory of relativity, but the velocity relative to the local standard of rest. It is quite clear that in all dynamical calculations, such as, for example, the advance of perihelion of a planet, the local standard of rest—the standard fixed by the extragalactic nebulae as a whole—is destined to play a much more important part than that attributed to it in the theory of relativity.

If, on the other hand, an extragalactic observer O is chosen so that $\boldsymbol{V} = 0$, then the expression reduces to $m/(1 - \boldsymbol{P}^2/c^2t^2)^{1/2}$ and measures the "potential" energy of the now ex-centric particle.

So much for the -1 in formula (4). What of the other term? This has an equally exciting interpretation. Since

$$(\xi - 1)^{3/2} = \frac{1}{c^3}\left[(\boldsymbol{P} - \boldsymbol{V}t)^2 - \frac{(\boldsymbol{P}\wedge\boldsymbol{V})^2}{c^2}\right]X^{-3/2}Y^{-3/2},$$

the acceleration term in (2) corresponding to the second term in (4) reduces, under well-defined circumstances, to

$$\frac{c^3C}{\psi(1)}\,\frac{\boldsymbol{P} - \boldsymbol{V}t}{(\boldsymbol{P} - \boldsymbol{V}t)^3},$$

which is an inverse-square attraction toward the point $\boldsymbol{V}t$, which is the center of the neighboring condensation in the statistical system. The essential point is that, without at any point making any assumption corresponding to the inverse-square law, an inverse-square law has emerged. It is, of course, the law of local gravitation. The curious index $3/2$ leads precisely to the negative exponent 2 in the distance of the particle from the condensation. It must be empha-

sized that the arguments employed in deducing (4) are purely kinematic; yet, an inverse-square law has appeared.

By appropriate analysis, this formula can be modified to give the gravitational potential χ for any pair of particles in the form

$$\chi = -\frac{m_1 m_2 c^2}{M_0} \frac{t_1 t_2 - \frac{\boldsymbol{P}_1 \cdot \boldsymbol{P}_2}{c^2}}{\left[\left(t_1 t_2 - \frac{\boldsymbol{P}_1 \cdot \boldsymbol{P}_2}{c^2}\right)^2 - \left(t_1^2 - \frac{\boldsymbol{P}_1^2}{c^2}\right)\left(t_2^2 - \frac{\boldsymbol{P}_2^2}{c^2}\right)\right]^{1/2}}, \quad (5)$$

where M_0 is the mass of the equivalent homogeneous universe and t_1 and t_2 are the epochs at the two particles. Here we have introduced the masses m_1 and m_2 of the two particles, and this leads us to refer for a moment to the arbitrary constant C in formula (4).

Why should this formula contain an arbitrary constant? Given the function $\psi(\xi)$, determining the population of the statistical system, it appears that there yet remains a degree of arbitrariness in the acceleration of a free particle. Why is this? The answer is that C corresponds to the *mass* of the condensation; C is thrown up by purely kinematic arguments, which are thus capable of predicting deductively the existence of gravitational mass. The actual connections are given by

$$M_0 = \tfrac{4}{3}\pi m_0 B ,$$

$$m_2 = \frac{C}{\psi(1)} M_0 ,$$

where m_0 is the arbitrary mass number attached to a fundamental particle. The details are of less interest than the unexpected fact that purely kinematic arguments insist on the appearance of a new parameter which can be identified with what is known empirically as "gravitational mass." This gives a kinematic basis to gravitation. We are ultimately led to the relation $\gamma = c^3 t / M_0$, which allows a determination of M_0 from t, the age of the universe, and γ, the "constant" of gravitation.

Thus, the kinematic arguments leading to (4) establish: (1) the law of inertia for a substratum; (2) the existence of "gravitational" mass; (3) the inverse-square character of "gravitational" force. These are thus removed from the domain of empirical knowledge to

the domain of deductive knowledge, though, of course, the appeal to observation remains in the act of *identifying* in nature entities corresponding to those thrown up deductively.

I have not time here to show in detail how these ideas can be enlarged so as to include electromagnetic forces. It suffices to remark that a "gravitational force" is derived by the differentiation of the potential (5), which is the simplest scalar potential which can be built up possessing a singularity at $\boldsymbol{P}_1 = \boldsymbol{P}_2, t_1 = t_2$. That is to say, a gravitational force is the simplest force, arising in a substratum, that can "act" on a free particle. If we inquire as to the next simplest type of "force," it appears that it arises from the double differentiation of super potentials, scalars which also possess singularities at $\boldsymbol{P}_1 = \boldsymbol{P}_2$ and have certain other properties. These "next simplest" types of forces then prove to be identifiable with electromagnetic forces. The further study of these, as between neighboring "charged" particles, leads to results of some interest in the theory of nuclear dynamics. It may be remarked that, just as "gravitational mass" is simply the numerical label distinguishing the simplest type of singularity possible, so "charge" is the numerical label distinguishing the next simplest type of singularity.

Much of what I have been saying is not in accordance with certain current ways of thinking. But it appears to me that the experimental and observational method—marvelous, as it is, for discovery—is powerless to indicate the answers to the deeper philosophical questions of the origin of laws of nature—to tell us why they hold good. This requires a purely logical analysis. Actually, even a logical analysis cannot tell us why they hold good, any more than a geometry can tell us why theorems hold good. But just as a geometry can tell us what geometrical theorems hold good, and give us a degree of belief in their validity deeper than can be gained from the principle of induction, so kinematical analysis (geometry and time) can tell us *what* kinematical, dynamical, gravitational, and electromagnetic theorems must hold good. For example, an appeal to the principle of induction cannot distinguish between an inverse-square law $1/r^2$ and a law $1/r^{2+10^{-9}}$; and the so-called "appeals to simplicity" urged by Jeffreys carry little conviction. Kinematic analysis leads to the definite law $1/r^2$, in its relativistic form, for local gravitation, thus de-

riving a so-called "law of nature" purely from the self-consistency of all the elements of observation with which that law deals. The details of this analysis are unlike the usual type of physical argument, and their very unfamiliarity tends to make them suspect. I am content to let the march of time bring that conviction of their reliability which has come to all who have studied them without preconceived prejudices.

In conclusion, I may mention that kinematic analysis gives definite answers to questions which can be asked about the universe but to which empirical relativity gives ambiguous answers. Among these are the questions:

1. Is the universe expanding or nonexpanding?
2. Is it finite in spatial extent or infinite?
3. Is the number of particles in it finite or infinite?
4. Is the space appropriate to it curved or flat?
5. Is its past history finite?
6. Is the universe homogeneous or not?

Kinematic analysis says that the answers to these questions depend on the scale of time adopted and that the adoption of a scale of time is an arbitrary act on the part of the thinker. Thus, if the scale of time is that in which the laws of dynamics take their classical form (τ-scale), then the universe is not expanding, is infinite in spatial extent with an infinite past history, and the appropriate public space is hyperbolic. If the scale of time is that in which an atomic frequency is taken as constant (t-scale), then the universe is expanding uniformly, is finite in spatial extent with a finite past history, and the appropriate private spaces are flat. In the former case the density distribution is homogeneous; in the latter case the density increases outward from the observer—and, indeed, from every observer. In both cases the number of particles is infinite, but there is a parameter M_0 which plays the part of a finite mass for the universe.

These answers depend fundamentally on only one axiom, namely, that a meaning can be attached to saying that different observers in different parts of the (expanding or nonexpanding) universe can set up the same measures of time when a precise arithmetical meaning

is given to this phrase. The element of experience incorporated into the analysis is the individual observer's consciousness of the passage of time, i.e., his awareness of a before-and-after relation between any pair of events he experiences. Starting from this awareness, and an initial arbitrary graduation of temporal experience, it proves possible to isolate that measure of time, or graduation of temporal experience, in which a free particle, projected among a "homogeneous" set of equivalent time-keepers, conserves its velocity. In this measure of time the system is stationary, nonexpanding, and of infinite extent. The red shift is, therefore, evidence that atomic vibrations are not conserved in period in this particular measure of time. Further analysis shows that they must be conserved in a (in some ways) more fundamental measure of time, that in which to any particular observer the universe is expanding, finite in extent but still infinite in population, and nonhomogeneous, with density increasing outward from each individual observer. The physics appropriate to this measure of time contains no arbitrary constant but only conventional constants like c and M_0. The theorems of the analysis are recognizable as laws of nature, but from their logical origin they have a sanction exceeding that of any law of nature depending on the principle of induction and on the supposed principle of the uniformity of nature. The latter principle stands exploded, for in a universe which is a seat of change, where epoch has a meaning, the laws of nature themselves must be functions of epoch, reckoned from some natural zero. The spectator of the universe can take his choice. If he regards it as changing, epoch is significant; and the laws of nature will, in general, be expected to involve this epoch. If it is stationary, then dependence on epoch disappears; but the existence of the red shift then shows that atomic emission and absorption frequencies are not constant in the time scale chosen. The existence of the two time scales is closely connected with the disjunction between matter and light. The τ-scale is appropriate to material particles and their dynamics and electrodynamics; the t-scale, to radiation. It is tempting to say, in conclusion, that it is the interplay of these two time scales which gives the universe its interest.

WADHAM COLLEGE
OXFORD, ENGLAND

SPECTRAL ENERGY-CURVE OF THE SUN IN THE ULTRAVIOLET*

EDISON PETTIT

ABSTRACT

1. *Apparatus.*—The Mount Wilson observations were made with the double quartz monochromator (slightly modified) that was used in the Tucson work. Light was supplied by an aluminized siderostat mirror. Both lens-projection and mirror-projection systems were used.

2. *Measurements.*—The measurements were made during the years 1934, 1937, and 1939, with the thermopile, by the method of constant dispersion. The results, reduced to no atmosphere, yield spectral energy-curves for integrated sunlight and for light from the center of the disk which verify the principal features obtained at Tucson, namely, a sudden drop of 48 per cent in intensity from λ 0.40 to λ 0.38 μ, a nearly constant intensity from λ 0.38 to λ 0.325 μ, and a nearly linear fall of intensity to a very low value at λ 0.292 μ.

3. *Atmospheric transmission coefficients.*—The means of determinations on 35 days, which include both integrated light and light from the center of the disk, do not differ more than 2 per cent from those of the Smithsonian observers until the ozone band is reached. Corrections due to the fall of energy within the second slit have been applied to the results for λ 0.295 and λ 0.292 μ.

4. *Atmospheric ozone content.*—A plot of atmospheric transmission against wave length shows that the ozone content is about 0.10 cm at N.T.P., only slightly more than half that found at Tucson.

5. *Energy between the lines.*—To eliminate as far as possible the effects of Fraunhofer line absorption, energy-curves were obtained with a 21-foot concave-grating monochromator and photoelectric cell over 100 A units of the spectrum. The areas below contour lines drawn over the highest ordinates, divided by the areas below the record of the line spectrum, furnish coefficients by which the thermoelectric energy-curves may be reduced to an energy-curve for the continuous spectrum. Although this procedure smooths out the ultraviolet energy-curve to some extent, most of the major depressions remain, and the curve in the ultraviolet is still far removed from that of a black body.

6. *Band spectrum of ozone.*—An earlier laboratory investigation of the ozone spectrum is reported. A calibration of the plates shows that the line intensities are so low that they will not be distinguishable in the solar spectrum and cannot alone account for depressions found in the photoelectric energy-curves.

7. *Line-group intensities.*—From the photoelectric records of the 21-foot monochromator the line-group intensities for the regions λλ 3200–3300 A and λλ 3900–4000 A are given. A comparison with Mulders' calibration of Rowland intensities in the latter region shows an average ratio of 0.81, with considerable scattering.

A determination of the distribution of energy in the ultraviolet solar spectrum made at Tucson, Arizona (elevation 2600 feet), in 1931, has been described in a previous paper.[1] As most of this work concerned the spectrum of *integrated* sunlight, it was thought useful to repeat the observations on Mount Wilson with special emphasis

* *Contributions from the Mount Wilson Observatory, Carnegie Institution of Washington*, No. 622.

[1] *Mt. W. Contr.*, No. 445; *Ap. J.*, **75**, 185, 1932.

on the radiation from the center of the disk, which, after all, best represents what the sun emits from the photosphere.

The observations at Tucson were made with a double Bausch and Lomb quartz monochromator and a vacuum thermopile. Sunlight was supplied with a Foucault underdrive siderostat with a 12-inch stellite mirror. The image-forming lens was of fused quartz, of 12 inches aperture and 6 feet focal length.

In arranging the equipment for operation on Mount Wilson, the same monochromator and thermopile were used, but either a 6-inch fused quartz lens of 6 feet focal length or a mirror system replaced the lens used at Tucson.

CONCERNING SIDEROSTATS

Siderostats are of two types. That used at Tucson (Fig. 1, *A*) has the mirror set perpendicular to the driving rod *R;* the declination arm *D* connecting it swings in an arc beneath the mirror. This arrangement has the disadvantage that in the original adjustment of the instrument it is not easy to make the horizontal beam of light *MN*, projected back through the mirror, intersect the polar axis at the same point *O* with the declination arm. In the overdrive type (Fig. 1, *B*) adopted at Mount Wilson this difficulty is overcome by drilling the axis *a* of the mirror fork *f*, which now points in the beam direction *mn*. The ends of the drilled axis *a* are closed with glass plates *pp'* having the centers marked. These act as sights. In setting up the siderostat for the initial run, the mirror is removed, and the axis sighted upon the optical part to receive the light. The mirror is then replaced, and the declination arm *d*, provided with a lens finder *ss'*, having been set upon the sun, is changed slightly in length until the beam from the mirror falls centrally upon the optical part desired. This procedure secures the most accurate following. The same roller slip-connection *k* between the declination arm *d* and the mirror driving rod *r* (now in the plane of the mirror) that was devised for Tucson has been adopted in the new instrument with equal success. An elevating screw *e* and an azimuth screw *h* facilitate the adjustment of the projected beam *mn* in both azimuth and altitude; the latter adjustment is mechanically difficult to effect with type *A*. The attachment *z* is an air-mass meter.

This siderostat, provided with a 12-inch aluminized mirror, was made part of the remounted 20-inch telescope and has electrical controls in both co-ordinates. The telescope, mounted on rails, can be

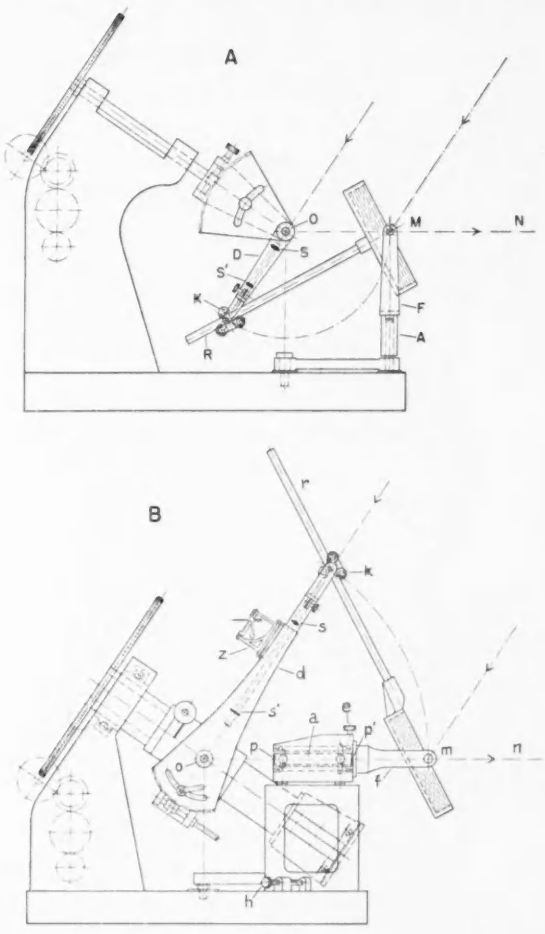


FIG. 1.—*A*, Underdrive siderostat as used at Tucson in the earlier measurement of the solar energy-curve in the ultraviolet. *B*, Overdrive siderostat as used on Mount Wilson.

moved in the meridian a distance of 8 feet. The beam from the siderostat enters the building to the south where the double quartz monochromator and a 21-foot concave-grating monochromator are placed.

THE DOUBLE QUARTZ MONOCHROMATOR

This consists of two monochromators with a common straight slit, the first and third slits being curved. The width of the second slit, which determines the wave-length range of the radiation measured in a given setting, was read on a 5-inch dial graduated to 100 parts. The four focusing pinions are now operated by a common knurled head.

The procedure in making the measurements, with some modifications, is the same as that already described.[2] In half the work of 1934 and in all that of 1937 an all-mirror projection system replaced the quartz lens, which eliminated the focusing of this part of the equipment for each wave length. This system involved two or three reflections from aluminized surfaces. In reducing the 1937 observations it was found that, possibly owing to prolonged exposure to sunlight, the reflecting power of the mirrors to the violet of λ 0.37 μ had fallen sharply. In 1939 this mirror system was therefore abandoned in favor of the single mirror and fused quartz lens, an arrangement used at Tucson in 1931 and on Mount Wilson during a part of 1934. The principle of eliminating all unnecessary reflecting surfaces in ultraviolet radiometric determinations seems to be well established by this experience.

In beginning the observations on Mount Wilson it was found that, owing to the peculiar construction of the slits, the changes of temperature in the early summer seriously affected the slit widths. This was especially troublesome in the case of the first slit, since the zero of the second was easily tested. At Tucson the temperature range was small during the observing season and therefore did not produce much interference. In 1934 and 1937 the temperature of the observing room on Mount Wilson was in part controlled, and correcting formulae were devised to eliminate, in so far as possible, the outstanding effects. In 1939 the first slit was replaced by a fixed invar slit. The third slit was also removed, and the thermopile placed directly in the spectrum. The fused quartz transfer lens formerly used was thus eliminated. No photoelectric observations were made as in previous years, but the system, more sensitive than

[2] *Mt. W. Contr.*, No. 445, p. 26; *Ap. J.*, **75**, 210, 1932.

that employed at Tucson, required a 5 per cent sector instead of the 10 per cent sector previously used, thereby making the deflections in the extreme ultraviolet, when the sector was stopped, double those formerly observed.

The effective length of the first slit was fixed by the length of the thermopile receivers, 4 mm. In 1934 and 1938 the diameter of the solar image was 16.8 mm so that the slit integrated the radiation over a region extending out from the center on either side to a distance 0.24 of the solar radius. Since the darkening along this fraction of a radius is small for all wave lengths, the measurements of spectral energy refer effectively to the center of the disk. In 1937 a solar diameter of 50 mm was used.

THE OBSERVATIONS

The program of observations was the same as that followed at Tucson. Since there were no considerable sky obstructions, the measures began early, when sec $z = 4$ to 7. To enable the observer to read the air mass at any time, the air-mass meter[3] was removed from the lunar equipment and attached to the declination arm of the siderostat (z, Fig. 1, B). This arrangement assists in securing the best distribution of points in the plots of the logarithms of deflections against sec z.

The method of constant dispersion was used throughout. This procedure involved the following adjustments of the double monochromator: The first slit was fixed at 1 division opening (0.1 mm, equal to 10 divisions on the second slit head); it is important that this opening remain undisturbed. The second slit was set for an opening which allowed 100 A of the spectrum of the first monochromator to pass through. The third slit was set to the same reading as the second plus one division. In the earlier work its image was transferred to the thermopile with a quartz lens, but in 1939, as already noted, both slit and lens were eliminated, and the pile placed directly in the image formed by the second monochromator. The method of constant dispersion has the advantage that the deflections of the galvanometer, after correction for transmission, reflectivity, etc., are directly a measure of the spectral energy. The method of constant slit

[3] E. Pettit, *Mt. W. Contr.*, No. 504; *Ap. J.*, **81**, 17, 1935.

widths, on the other hand, involves dispersion corrections, which in the extreme ultraviolet become very large, more than 600 per cent at λ 0.30 μ with a quartz monochromator.

The wave lengths in which intensities were measured were the same as those used at Tucson and are indicated in Table 1. Each measurement consisted in the following operations: (1) The two wave-length dials of the monochromator were set. (2) With the aid of a table (second col., Table 1) on the observing sheet, the second slit was set to transmit 100 A. (3) Until its removal in 1939, the third slit was set to this value plus 1 division. (4) The monochromators were focused. (5) The quartz projecting lens was focused. (6) A guiding image from the siderostat mirror was checked. (7) After the shutter was opened, the wave-length dial of the first monochromator was moved slightly by a lever slow-motion until maximum deflection was observed. (8) The hour angle was noted. All these operations were repeated for each of the 22 wave lengths. Further, the temperature and second-slit zero reading were noted at the beginning and end of each set.

One such single set of measurements required about a half-hour to complete. The first two sets were run in immediate succession as soon as the observations began after sunrise. An interval of a half-hour was allowed between the second and third and between the third and fourth sets. A fifth and sometimes a sixth set was taken during the hour just preceding meridian passage. Usually only the wave lengths in the extreme ultraviolet were observed in the last set, since the intensities of the higher wave lengths showed little change.

In all these measurements a rotating sector transmitting 5 per cent was employed until the deflections dropped to about 25 mm; then the sector was stopped, with the opening in front of the slit, thus increasing the deflections to 500 mm. This usually occurred near λ 0.33 μ for the first set, but near λ 0.30 μ in the last. No deflections could be found at λ 0.30 μ in the first and at λ 0.295 and λ 0.292 μ in the first two sets. This is illustrated in Table 1. It was customary therefore to add a few measures in this region as the sun approached meridian. In the work on integrated sunlight done in 1934 this region was measured with the sodium photoelectric cell,

with an overlap to connect with the thermoelectric measures in a longer wave length.

As an illustration, Table 1 gives the observing record for June 8, 1939. Besides λ and the second-slit setting S_2, the table gives the

TABLE 1

SOLAR ENERGY-CURVE OBSERVATIONS, JUNE 8, 1939
(Morning)

λ	S_2	Set 1 S_2 Zero = +0.05		Set 2 S_2 Zero = +0.05		Set 3 S_2 Zero = +0.06		Set 4 S_2 Zero = +0.06		Set 5 S_2 Zero = +0.06	
		T = 22°.5 C		T = 23°.5 C		T = 25°.0 C		T = 26°.3 C			
μ			mm		mm		mm		mm		
0.70	0.43	6h13m	207	5h40m	250	5h03m	267	3h00m	288	.....	
.60	0.70	6 11	281	5 38	358	5 02	408	2 59	442	.....	
.55	0.89	6 10	271	5 36	371	5 00	428	2 57	491	.....	
.50	1.17	6 09	273	5 35	372	4 58	448	2 56	513	....	mm
.45	1.61	6 08	245	5 34	366	4 56	445	2 54	546	0h23m	562
.42	2.00	6 06	154	5 32	250	4 55	317	2 53	407	0 22	429
.40	2.34	6 05	128	5 31	224	4 53	305	2 52	423	0 20	426
.39	2.54	6 03	96	5 28	174	4 51*a*†	240	2 50	323	0 20	332
.38	2.76	6 02	57	5 27	112	4 08*b*	188	2 49	224	0 18*c*	245
.37	3.01	6 00	50	5 25	105	4 05	189	2 47	228	0 17	249
.36	3.30	5 59	41	5 24	91	4 04	172	2 46	210	0 15	228
.35	3.63	5 58	*x*600*	5 23	68	4 03	139	2 45	174	0 14	195
.34	4.02	5 56	*x*451	5 22	54	4 02	121	2 44	155	0 13	175
.33	4.45	5 55	*x*315	5 20	41	4 01	101	2 42	138	0 11	156
.325	4.69	5 54	*x*239	5 19	*x*670	4 00	85	2 40	118	0 10*d*	132
.32	4.90	5 52	*x*141	5 17	*x*430	3 58	59	2 39	86	1 28*e*	104
.315	5.20	5 50	*x* 60	5 15	*x*227	3 57	37	2 38	59	1 26	70
.31	5.47	5 49	*x* 22	5 13	*x* 97	3 55	*x*414	2 36	35	1 25	42
.305	5.75	5 47	*x* 4	5 12	*x* 25	3 54	*x*151	2 35	*x*301	1 24	*x*395
.30	6.05	.....		5 11	*x* 3	3 53	*x* 32	2 34	*x* 87	1 22	*x*125
.295	6.37	.....		.....		3 50	*x*2.3	2 32	*x* 9	1 20	*x* 14
0.292	6.57	.....		.....		3 48	*x*0.5	2 30	*x*1.5	1 18	*x*3.0
		T = 23°.5 C		T = 24°.8 C		T = 26°.0 C		T = 27°.4 C		T = 27°.9 C	
Range in sec z	.....	5.76–3.86		3.52–2.64		2.47–1.58		1.32–1.22		1.02–1.08	

* The symbol *x* indicates sector was stopped, increasing deflections twenty fold.

† Additional temperature observations: *a* = 25°.3, *b* = 25°.5, *c* = 28°.5, *d* = 28°.8, and *e* = 27°.6. In set 5 the measures to the violet of λ 0.325 μ preceded the others.

data for five sets of measures, consisting of the hour angle followed by the observed galvanometer deflection in millimeters. The notation *x* before a deflection means that the 5 per cent sector was stopped with its aperture in front of the slit, thus multiplying the

deflections by 20. At the top of each column is noted the second-slit zero and below it the temperature. The temperature is also noted where a discontinuity occurs in the hour-angle column and at the end of each set. Below is given the range in sec z (air mass) for that column. The negative sign of the hour angle has been omitted, as it applies throughout the table.

TEMPERATURE EFFECTS

There are several possible effects of temperature changes on the thermocouple and the galvanometer circuit. The resistance of most metals, particularly those used in making the thermopile and the copper of the galvanometer circuit, increases about 0.4 per cent per degree centigrade. On this account alone we would expect the deflections to diminish by a like amount with increasing temperature. There is also the slightly decreasing torsion of the galvanometer suspension, which would increase the deflections, and the slight change in thermoelectric power of the thermopile and its decreasing heat engine efficiency, $T^{-1}\Delta T$. Against these considerations must be placed the uncertain effect of temperature on the vacuum in the thermopile cell with its calcium tube. This is a very important item, for, if the vacuum is near the limit corresponding to the steep slope of the pressure-sensitivity curve,[4] the deflections may be considerably decreased if the temperature change slightly raises the pressure.

To test the question, the thermopile (freshly exhausted) and the galvanometer used in the present measurements were inclosed in a paper box, provided with windows to admit light to the thermopile (one receiver of which was shaded) and to expose the galvanometer mirror. The temperature was controlled by a 75-watt lamp within the box. The deflections were obtained by switching on a 100-watt lamp at 3 meters' distance. This lamp, with a voltmeter across the circuit, was fed from a D.C. generator. The voltage was maintained at 115 by means of a microvoltage control. The readings made are given in the following table. For the first 10° of temperature increase the deflection decreased only 1 per cent, and as this interval includes the range usually occurring during observations of the solar energy-curve, we may conclude that the effect of ordi-

[4] E. Pettit and S. B. Nicholson, *Mt. W. Contr.*, No. 246, p. 7; *Ap. J.*, **56**, 301, 1922.

nary temperature changes on radiation sensitivity may be neglected.

Temperature (C)	23°.8	27°.0	30°.4	33°.5	39°.0
Deflection (mm).....	339	338	338	335	320

The slit mechanisms were such that increasing temperature decreased the slit widths. While the effect on the first slit differed from that on the second and the third, we may for convenience suppose that the effect on all three is involved in the second slit alone. Let us suppose that the normal width of the slit is 0.043 mm for the first morning observation at λ 0.7 μ and the corresponding deflection D_0; succeeding settings give D_1. . . . . Since the atmospheric transmission on Mount Wilson for λ 0.7 μ is 0.95 (Abbot), the deflection D for any air mass is

$$\log D = (\log D_0 - 9.978 \sec z_0) + 9.978 \sec z. \qquad (1)$$

If the fundamental atmospheric transparency is constant and if the sensibility of the galvanometer circuit remains unchanged or can be corrected, we may attribute any deviation of the observed deflections from those calculated by equation (1) to change of slit width. Suppose the observed deflection is D', then the change in slit width for any sec z' is

$$\delta w = 0.043\ (D'D^{-1} - 1)\,. \qquad (2)$$

In practice the values of δw were plotted against hour angle and the corrections to slit width were taken from the curve. These corrections apply to all wave lengths. The correct deflection D_c for any wave length at any air mass is then

$$D_c = D_a w(w + \delta w)^{-1}\,, \qquad (3)$$

where w is the tabular slit width and D_a the observed deflection. The value of δw did not usually exceed $\pm$0.008 mm.

PROPORTIONALITY OF DEFLECTIONS TO SLIT WIDTH

Since the second slit is opened in proportion to the dispersion for each wave-length setting, it is important to know if the deflections

TABLE 2

Spectral-Energy Measurements

λμ	S_2	S	K_1	K_2	K_3	K_4	K_5	1931 *i*	1934 *i*	1931 *c*	1934 *c*	1937 *c*	1939 *c*	Av. *i*	Av. *c*	Tra.
0.70	0.43	1.38	0.93	0.95	0.92	0.92	0.96	11.0	16.1	14.4	16.8	16.1	14.6	13.6	15.5	0.943
.60	0.70	1.21	0.93	0.95	0.92	0.87	0.97	18.3	21.1	22.4	23.1	24.0	21.7	19.7	22.7	.911
.55	0.89	1.09	0.96	0.96	0.94	0.92	0.98	20.6	20.9	24.0	23.3	24.6	22.7	20.8	23.6	.882
.50	1.17	1.07	0.98	0.99	0.97	0.96	0.99	22.4	22.0	25.9	24.8	26.8	24.6	22.2	25.5	.869
.45	1.61	1.00	1.00	1.00	1.00	1.00	1.00	22.4	22.4	26.0	26.0	26.0	26.0	22.4	26.0	.832
.42	2.00	0.98	1.01	1.00	1.02	1.02	1.00	17.7	17.9	22.0	19.9	21.4	22.0	17.8	21.3	.789
.40	2.34	0.96	1.03	1.01	1.03	1.04	1.01	17.5	18.0	20.9	19.7	20.5	22.0	17.7	20.7	.752
.39	2.54	0.95	1.04	1.01	1.05	1.06	1.01	12.9	14.1	12.5	15.3	14.7	19.3	13.4	15.5	.742
.38	2.76	0.94	1.05	1.02	1.06	1.08	1.02	10.3	10.4	11.5	11.4	10.8	13.8	10.3	11.9	.705
.37	3.01	0.93	1.06	1.02	1.07	1.11	1.02	11.3	11.0	12.3	11.6	12.9	13.0	11.2	12.4	.669
.36	3.30	0.92	1.07	1.02	1.08	1.13	1.02	11.2	10.9	12.5	11.4	12.2	13.0	11.0	12.3	.650
.35	3.63	0.92	1.08	1.03	1.10	1.15	1.03	9.5	10.0	12.1	10.5	10.8	12.0	9.7	11.3	.605
.34	4.02	0.91	1.09	1.03	1.11	1.17	1.03	9.7	9.9	11.5	9.9	11.0	11.4	9.9	11.0	.567
.33	4.45	0.91	1.11	1.04	1.13	1.18	1.03	10.0	9.5	11.6	9.9	10.9	11.0	9.7	10.9	.521
.325	4.69	0.91	1.12	1.04	1.15	1.20	1.04	9.9	9.5	11.6	9.9	10.6	10.3	9.6	10.6	.491
.32	4.90	0.91	1.13	1.05	1.16	1.22	1.04	9.4	8.8	8.4	9.9	9.3	9.0	9.2	9.2	.456
.315	5.20	0.90	1.14	1.06	1.18	1.24	1.05	7.7	7.4	6.7	7.1	7.4	7.7	7.6	7.3	.401
.31	5.47	0.90	1.16	1.07	1.20	1.25	1.06	6.9	6.6	5.8	6.0	6.5	6.0	6.7	6.2	.329
.305	5.75	0.90	1.17	1.08	1.22	1.28	1.07	5.5	5.4	4.3	4.9	5.2	4.6	5.4	4.7	.246
.30	6.05	0.90	1.19	1.10	1.24	1.30	1.08	7.3	3.9	3.1	3.0	3.2	3.1	5.6	3.1	.166
.295	6.37	0.90	1.21	1.11	1.26	1.32	1.09	5.0	2.1	0.7	1.9	3.0	1.3	3.6	1.8	.088
0.292	6.57	0.90	1.22	1.12	1.28	1.34	1.10	2.6	1.5	1.1	1.5	3.5	1.0	2.0	1.6	0.047

are proportional to this opening. This was investigated for λ 0.32 μ with both thermopile and photocell, and for λ 0.6 μ with the thermopile. Deflections were read for slit openings increasing by intervals of 0.05 of a screw revolution over the range used in observation. When plotted against slit width, these deflections form a straight line which does not quite pass through the origin, showing that the proportionality rule holds save for slits narrower than 0.05 revolutions. In these data the reading for zero slit width was carefully determined; beginning at 0.05 revolutions, the line representing the plot is therefore curved as it approaches the origin. To obtain the corrections a line is drawn parallel to the plot through the origin; the ratio of any ordinate on this line to that on the plot will be the correction to an observed deflection to make it proportional to slit opening. These corrections appear in Table 2, column *S*.

REDUCTION OF THE MEASUREMENTS

The logarithms of the observed deflections (log *d*), corrected if necessary by equation (3), were plotted against air mass (sec *z*), and the straight lines best representing the observations were extended to the zenith (air mass 1) and to no atmosphere (air mass 0). Figure 2 shows such a plot for a satisfactory day's run. From these plots the values of log d_0 and log d_1 were read for air masses 0 and 1, giving d_0, the deflection for no atmosphere, and log t_r = log d_1 − log d_0, the logarithm of atmospheric transmission. Inspection of Figure 2 shows that the values of d_0 between λ 0.38 and λ 0.325 μ cluster about the same ordinate on air mass 0. Since the corrections to be applied to d_0 are not large, it is evident that the ultraviolet solar intensity does not vary greatly over this range in wave length.

The corrections involved depend somewhat on the arrangement of the optical system. There is always the reflecting power of the siderostat mirror. During part of 1934 there was one, and throughout 1937 a second, additional mirror. For the remainder of 1934 a fused quartz lens replaced the added mirror; and this lens was also used in 1939. Then there are two quartz monochromators and, except in 1939, a fused quartz transfer lens. The window on the vacuum thermopile is so thin (1.5 mm) that the losses are sensibly those by reflection from the two surfaces. According to Fresnel's formula, the

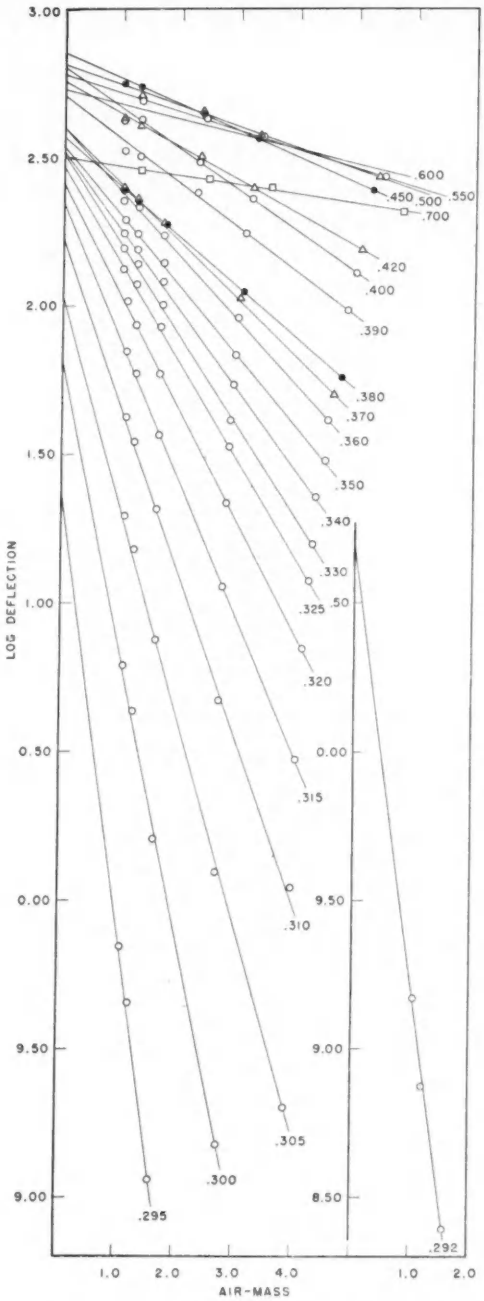


FIG. 2.—Plot of logarithm of deflections against air mass, for June 8, 1939. Note that between λ 0.38 and λ 0.325 μ the deflections at air mass 0 have approximately the same value.

loss for one surface at λ 0.3 μ is 0.0384 and at λ 0.7 μ, 0.0344—a range of less than 0.5 per cent, which can be neglected. When the photoelectric cell was used, a fused quartz prism was thrown in front of the transfer lens to send the light to the cell. Since the optical paths through the prism and lens are about the same, this alteration of conditions need not be considered.

If R_a be the reflecting power of an aluminized surface,[5] T_R the transmission of fused quartz[1] (both the image-forming and the transfer lenses were of approximately the same thickness), T_r the transmission of one monochromator,[1] F the focal length (or aperture ratio) of the fused quartz image-forming lens, and S the reciprocal of the slit efficiency, we have the following correction coefficients K to be applied to the mean value of d_0 for each wave length:

$$1934\,i \qquad K_1 = S\,R_a^{-1}\,T_r^{-2}\,T_R^{-1}\,, \tag{4}$$

$$1934\,A \qquad K_2 = S\,F^2\;R_a^{-1}\,T_r^{-2}\,T_R^{-2}\,, \tag{5}$$

$$1934\,B \qquad K_3 = S\,R_a^{-2}\,T_r^{-2}\,T_R^{-1}\,, \tag{6}$$

$$1937 \qquad K_4 = S\,R_a^{-3}\,T_r^{-2}\,T_R^{-1}\,, \tag{7}$$

$$1939 \qquad K_5 = S\,F^2\;R_a^{-1}\,T_r^{-2}\,T_R^{-1}\,. \tag{8}$$

Formula (4) refers to integrated light, (5) to a lens projector, (6) to a single-mirror projector, (7) to a two-mirror projector; (8) is the same as (5) with the transfer lens and third slit eliminated. The numerical values of K are given in Table 2. It will be noted that the ranges in K_2 and K_5 are much less than for the other coefficients. This is due to the compensating effect of the change in focal length F of the quartz projecting lens with wave length. In fact, if a telephoto arrangement were used, a nearly complete compensation would be possible.

After the averages were taken and before K was applied, the measurements were reduced to a scale of watts per square meter per 100 A by making the values at λ 0.45 μ equal to 22.4 for integrated light and to 26.0 for the center of the disk. The method of arriving at the former value is as follows: The energy in the integrated

[5] E. Pettit, *Pub. A.S.P.*, **46**, 27, 1934.

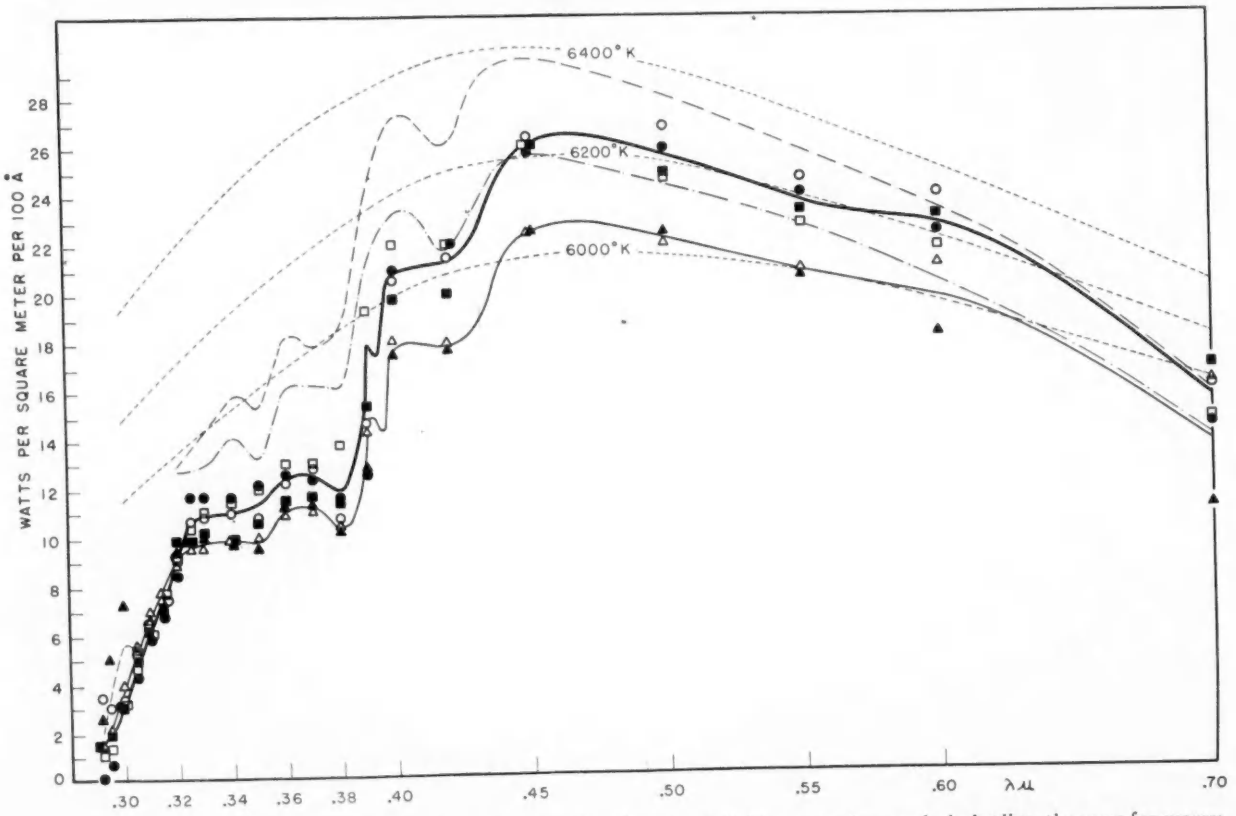


FIG. 3.—Solar energy-curve in the ultraviolet. Light line, integrated light 1931 and 1934; dash-dot line, the same for energy between the Fraunhofer lines; heavy line, center of the disk 1931, 1934, 1937, and 1939; dashed line, the same for energy between the Fraunhofer lines; dotted lines, black-body curves reduced to mean solar distance.

solar spectrum[6] between λ 0.45 and λ 0.70 μ is 0.71 cal cm^{-2} min^{-1}, or 4.95×10^5 erg cm^{-2} sec^{-1}. If we plot the solar energy-curve for no atmosphere as given in arbitrary units in the *Smithsonian Physical Tables*[6] over this spectral region, we find that the area within an interval of 100 A at λ 0.45 μ is 0.0453 of the whole. The corresponding energy is $0.0453 \times 4.95 \times 10^5$, or 22.4×10^3 erg cm^{-2} sec^{-1} $(100\ A)^{-1}$, or 22.4 watts per square meter per 100 A. In the previous calibration used for the work at Tucson, the approximate value 20 watts was used; but it is believed that the value here found will give a more accurate scale, and it has therefore been adopted. A calibration from Planck's radiation formula for a solar temperature of 6000° K, reduced to mean solar distance, yields 21.4 watts per square meter.

To obtain the calibration for the center of the disk the area of the solar energy-curve was increased by the factor 1.16, the ratio[7] of the average intensity of total radiation for the whole solar disk to the intensity at the center. The slight shift of the maximum to the violet, which the Smithsonian observations seem to show in the energy spectrum for the center of the disk, has been neglected; it is too small to be of importance here.

The results in watts per square meter per 100 A are shown in Table 2, where *i* following the date indicates integrated light, and *c* light from the center of the disk. "Av." is the average value, including the 1931 work at Tucson reduced to the new scale. "Tra." is the average atmospheric transmission for all the observations on Mount Wilson, a total of 35 determinations for each wave length.

Figure 3 shows a plot of the results for each year's work. In comparing them it should be remembered that they have unequal weight, the number of days of observations being different, as shown in the following table. But since all these different series of measures apply to different years, no weighting of normal places was made in taking the averages in Table 2 or in drawing the mean curves of Figure 3. The same general features are seen in this plot as were found in the 1931 plot, viz., a sudden drop of 48 per cent in the solar

[6] *Smithsonian Physical Tables* (8th rev. ed.), p. 608, 1933.

[7] E. Pettit and S. B. Nicholson, *Mt. W. Contr.*, No. 397; *Ap. J.*, **71**, 153, 1930.

energy between λ 0.40 and λ 0.38 μ and a nearly constant value of the solar energy from λ 0.38 to λ 0.325 μ.

	1931	1934	1937	1939
Center	2	10	9	10
Integrated	4	7	.....	.....

The ordinates of the curve for integrated light are nearly everywhere proportional to those for the center of the disk. This is particularly interesting in the ultraviolet, where the rise in intensity between λ 0.38 and λ 0.35 μ is revealed in both curves. Below λ 0.32 μ, the integrated-light curve runs even a little higher than that for the center of the disk. This is hard to understand, since it is supposed that darkening at the limb would tend to make it fall even farther below than for higher wave lengths. The same phenomenon was noted[1] in 1931. There is some indication here that darkening at the limb does not vary greatly throughout the ultraviolet region. The high values for integrated light found in 1931 for λλ 0.30, 0.295, and 0.292 μ were not verified in 1934 and may be erroneous. The curve through the averages for these wave lengths has been shown here as a dotted line, the full line including only the Mount Wilson values. The fall in intensity from λ 0.325 to λ 0.292 μ is practically linear. This interval is all within the ozone band, and, while the effect of the band should be eliminated by the method of reduction, the steepness of the transmission plots (Fig. 2) leads to considerable uncertainty. This is all the more patent when we consider the scattering of points below λ 0.30 μ in Figure 3.

There is, however, an effect which would make the measures below λ 0.30 μ too high. The second slit opening, being always 100 A, includes the region λλ 0.30–0.29 μ for the setting λ 0.295 μ, and λλ 0.297–0.287 μ for the setting λ 0.292 μ, and the thermopile integrates the radiation over these wave lengths. If the decrease of radiation across the 100-A slit decelerates toward the violet or if it falls to zero within the second slit, the observed deflections will be too high or will refer to a higher wave length than that corresponding to the center of the slit. For wave lengths greater than λ 0.30 μ

this criticism does not apply. This effect would also tend to make the atmospheric transmission coefficients too large.

ATMOSPHERIC TRANSMISSION AND OZONE

The atmospheric transmission coefficients have been found for both integrated light and light from the center of the disk, the data for

TABLE 3

ATMOSPHERIC TRANSMISSION ON MOUNT WILSON

$\lambda\mu$	1934	1937	1939	Av.	Smithsonian
0.7	0.950	0.938	0.941	0.943	0.950
.6	.905	.893	.943	.911	.900
.55	.884	.875	.888	.882	.....
.50	.869	.866	.871	.869	.862
.45	.827	.840	.827	.832	(.825)*
.42	.790	.784	.793	.789	.....
.40	.764	.748	.743	.752	.729
.39	.755	.740	.731	.742	.....
.38	.716	.699	.700	.705	.676
.37	.682	.655	.670	.669	.....
.36	.663	.640	.646	.650	.635
.35	.617	.607	.592	.605	.....
.34	.585	.565	.545	.565	.580
.33	.541	.520	.501	.521	.....
.325	.502	.495	.476	.491	.....
.32	.476	.462	.431	.456	.520
.315	.421	.403	.380	.401	.....
.31	.349	.328	.311	.329	.....
.305	.272	.238	.227	.246	.....
.30	.195	.167	.137	.166	(0.460)
.295	.114	.080	.071	.088†	.....
0.292	0.078	0.026	0.037	0.047†	.....

* Interpolated.

† When corrected for slit width, these become 0.071 and 0.032, respectively.

which may be combined by taking simple means for any one season. These coefficients were obtained for 15 days in 1934, 9 days in 1937, and 11 days in 1939. The averages for each year, together with the corresponding coefficients from the Smithsonian work on Mount Wilson,[6] are shown in Table 3.

We may determine the atmospheric ozone in the same manner as was done for the Tucson observations. From Rayleigh's law of molecular scattering,

$$\log t_r = -32\pi^3 (3N)^{-1} H \log e \, (\mu - 1)^2 \lambda^{-4}, \qquad (9)$$

we see that for zenith conditions, where the atmospheric depth H and the molecular density N are constant, the logarithm of the transmission, $\log t_r$, depends on the factor $-(\mu - 1)^2\lambda^{-4}$. Hence if we plot the observed $\log t_r$ against a wave-length scale which expands according to the function $-(\mu - 1)^2\lambda^{-4}$, we should expect a straight line as long as the atmospheric interference is of pure molecular scattering.

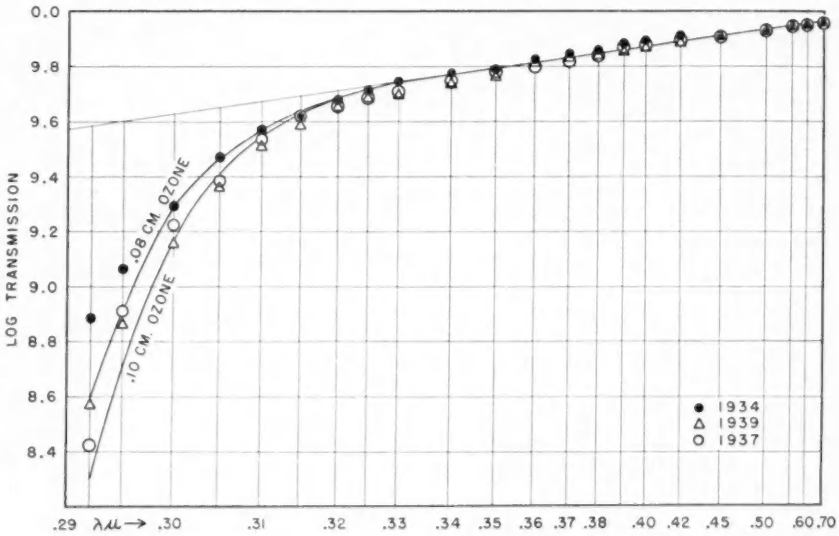


FIG. 4.—Atmospheric transmissions plotted on a wave-length scale that expands according to the function $-(\mu-1)^2\lambda^{-4}$. Heavy lines indicate ozone transmission for layers 0.08 and 0.10 cm equivalent thickness at N.T.P.

Figure 4 is such a plot of the observations for 1934, 1937, and 1939 from the data in Table 3. It is seen that the portion to the red of λ 0.33 μ is represented by a straight line, but for shorter wave lengths the deviation increases continuously. If we extend the straight portion of the line to the violet, the difference in ordinates for any wave length, in the sense of extended line *minus* observed curve, is the logarithm of the absorption due to atmospheric ozone.

The coefficient of absorption of ozone for a layer 1 cm thick at N.T.P. given by Fabry and Buisson is

$$\log a = 17.58 - 0.00564\,\lambda\,, \qquad (10)$$

where λ is in angstroms. The thickness of the equivalent ozone layer due to any observed atmospheric transmission t_R is

$$h = a^{-1} (\log t_r - \log t_R) , \qquad (11)$$

where h is in centimeters and $\log t_r$ is taken from the projected line for the same wave length.

The more direct way of estimating ozone is to compute ($\log t_r - \log t_R$) from equation (11) for assumed values of t and compare the curves with the plotted points. This has been done for $h = 0.08$ and 0.10 cm, whence it is seen that the Mount Wilson observations indicate an atmospheric ozone content of about 0.1 cm. The 1937 series gives 0.08, and the 1939 series about 0.11 cm. These values are considerably lower than that found at Tucson (0.18 cm) at the same season of the year.

It will be noted that the values of t_R for λ 0.295 and λ 0.292 μ are greater than those for the computed ozone-curves. The values in note † of Table 3 will bring these points between the two ozone-curves. This is again evidence that the measures in these wave lengths should be made with narrower slits.

On account of the low value of the atmospheric ozone content above Mount Wilson, it was thought that some error might have occurred in the calibration of the wave-length dials in 1934, which was done with a mercury arc before and after the series. This calibration was checked in the following winter, after the observations had been reduced, and no error was discovered. In 1937 the wave lengths were checked in the middle of the series, as well as before and after; and in 1939 the checks were made before and after each day's run. It seems then that the observed low value of the ozone content of the atmosphere above Mount Wilson cannot have its origin in wave-length error.

ULTRAVIOLET ENERGY BETWEEN THE LINES

The energy-curves in Figure 3 represent the total solar energy received from each interval of 100 A as modified by the absorption of the Fraunhofer lines. If these lines could be eliminated by measuring the radiation between them and if the wings of the lines did not seriously overlap, the result would be the energy-curve of the con-

tinuous spectrum of the sun. In the ultraviolet, however, there is no space between the lines in which the overlap of the wings can be ignored, and any direct measurement must give something less than the energy of the continuous spectrum.

To investigate the matter the 21-foot concave-grating monochromator, arranged in Eagle mounting, was provided with an 8-inch grating giving a bright first-order spectrum. The turntable supporting the grating is mechanically connected with the gear train that drives the recording drum of the galvanometer. Light entering the first slit and falling upon the grating returns as a spectrum along approximately the same path and passes through a second slit behind which is placed a vacuum chamber containing the quartz photoelectric cell and amplifying ("pliotron") tube. A fused quartz window admits light to the vacuum chamber. The control box and batteries are placed near the galvanometer (Leeds and Northrup 900 ohms). A 2-inch concave mirror of 26 feet radius is mounted on the grating vertical axis. This mirror casts an image of a straight-filament lamp upon a scale graduated in 100-A units for setting purposes. A fused quartz convex cylindrical lens can be put behind the first slit to bring the focal plane to the slit jaws for any wave length.[8] A filter holder is inserted in the light path near this lens.

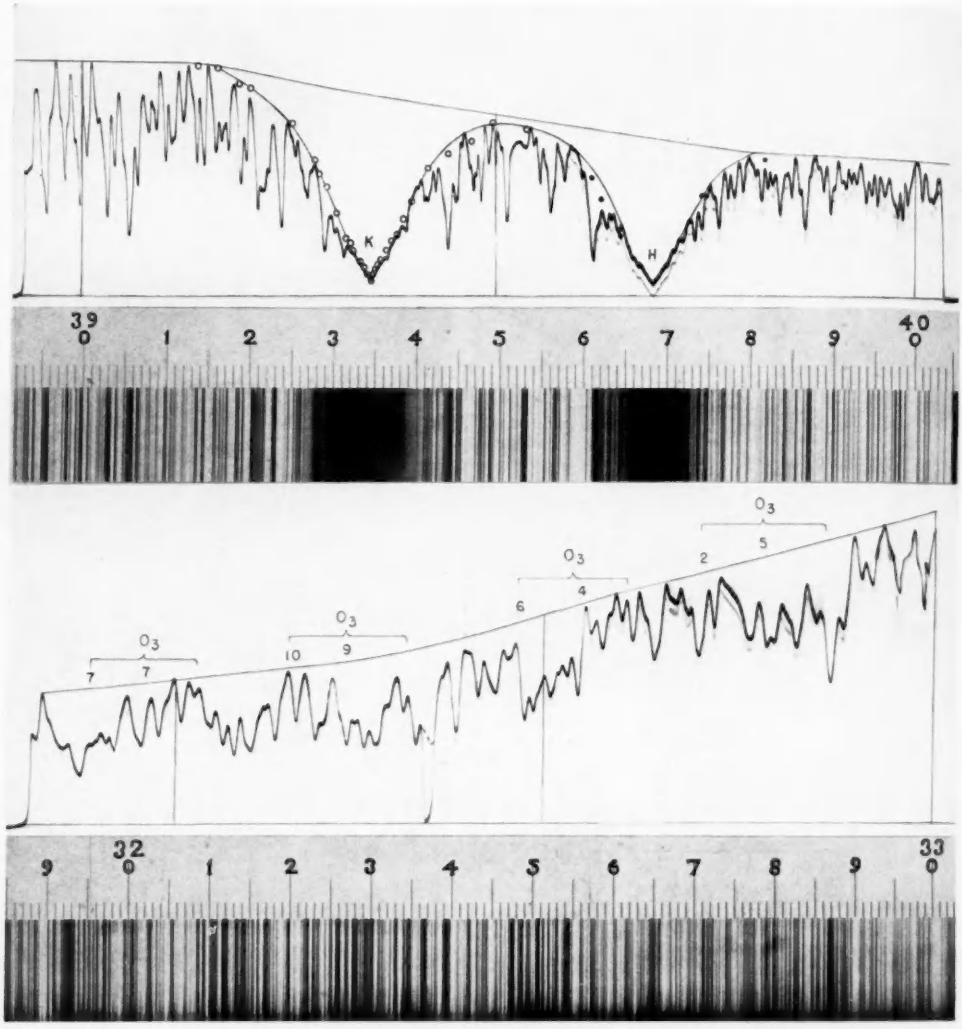
A concave-mirror projecting system of 40 feet focal length was used to form the solar image upon the first slit. Ordinarily the two slits had the same width, about 0.08 mm. For registration to the violet of λ 0.40 μ a Shott UG1 filter was used to reduce scattered light. Because of the sluggishness of the electrical system and the 3-second period of the galvanometer, about 15 minutes were required to record each 100-A interval. For this reason records were not made when sec z was greater than 2.

For the region $\lambda\lambda$ 0.49–0.37 μ the gas-filled potassium cell was used, but for shorter wave lengths the sodium cell was necessary. The color-sensitivity function of the sodium cell has already been given.[9] Tests made with a photographic step wedge showed that the galvanometer deflections were proportional to the radiation in the 21-foot monochromator when used with these cells.

[8] E. Pettit, *Pub. A.S.P.*, **43**, 75, 1931.

[9] *Mt. W. Contr.*, No. 445, p. 25; *Ap. J.*, **75**, 209, Table VII, 1932.

PLATE I



PHOTOELECTRIC ENERGY-CURVES FOR THE H AND K REGION (*above*) AND THE REGION λλ 3200–3300 A (*below*) COMPARED WITH ROWLAND'S MAP

The points on the contour of K are from Thackeray. Owing to faulty driving, λ 3053 is displaced 2A to the violet. Lines within 3A on either side are affected by smaller amounts. The braces O_3 indicate the positions of the principal ozone doublets and the numbers, the positions, and the relative intensities of their components. (The ordinate drawn at λ 3205 should be at λ 3200.)

Usually the recordings were made over the integral hundreds of angstroms, for example, λλ 0.35–0.36 μ, λλ 0.36–0.37 μ, etc. Contour lines were drawn through the apparently highest points of the records and subdivided into 50-A units (Pl. I). The areas below the contour line and below the curve itself were then measured with a planimeter. The ratio of the former to the latter, averaged for two contiguous regions, is the ratio of the energy in the 100-A interval, as it would be if measured between the lines most widely separated, to that measured with the thermopile and quartz monochromator shown in Figure 3. The data for the 50-A units were then combined by simple averages to give the ratios for λλ 0.355–0.365 μ, etc., which correspond to settings of the quartz monochromator at λ 0.360 μ, etc. This procedure was necessary since the planimeter was not large enough to measure an entire 100-A record.

Plate I shows galvanometer records for the familiar H and K region and for λλ 0.32–0.33 μ. In the former the contour has been drawn over the H and K lines for comparison with the contours found by A. D. Thackeray,[10] who used the method of photographic photometry with a 21-foot concave grating. Thackeray's observational points for his contour of the K line are shown by the heavy dots in the figure. Since he gives no data for the H line, some of his observational points for K have been transposed upon the H contour to show the agreement between the two contours; they should not be mistaken for observations of the H line itself. This comparison shows that the agreement for K is as close as could be expected from two such different methods of procedure and that the contours of H and K are nearly alike. The region λλ 0.32–0.33 μ is at the beginning of the ozone band, and the positions of several of the broad ozone lines are indicated on the curve by the braces under the symbol O_3.

Table 4 shows the ratio, e, of the energy between the lines to that in the integrated spectrum averaged over 100 A, found by the foregoing method. The column eM is obtained from estimates made by G. F. W. Mulders[11] based on a calibration-curve of Rowland's scale of intensities. The most serious disagreement is for λ 0.39 and λ 0.40 μ; but this at least cannot be due to scattered light in the

[10] *M.N.*, **95**, 293, 1935.

[11] *Zs. f. Ap.*, **11**, 132, Table 2, 1935.

21-foot monochromator, since it is near the red cutoff of the filter and the contour of the K line agrees with that of Thackeray, as shown in Plate I.

The quantities e Av. i and e Av. c are plotted in Figure 3 as broken lines. It is seen that the effect of correcting the thermoelectric measures for line absorption in this manner is to smooth out the violet slope of the energy-curve of the sun. Although the major features are still present, the energy in the ultraviolet no longer approximates a constant value. The black-body radiation-curves for

TABLE 4

RATIO e OF THE ENERGY BETWEEN THE LINES TO THAT IN THE INTEGRATED SPECTRUM

$\lambda\mu$	e	eM	e Av. i	e Av. c
0.320	1.39	1.23	12.8	12.8
.330	1.32	1.32	12.9	14.3
.340	1.43	1.33	14.1	15.7
.350	1.35	1.33	13.1	15.2
.360	1.47	1.54	16.1	18.1
.370	1.43	1.61	16.0	17.8
.380	1.56	1.59	16.1	18.5
.390	1.56	1.85	20.9	24.1
.395	1.61	.....	25.1	29.0
.400	1.31	1.54	23.2	27.1
.420	.....	1.23	21.7	26.0
.445	1.14	1.12	25.5	29.6
.485	1.12	1.11	24.8	28.6
.500	.....	1.10	24.4	28.1
.550	.....	1.09	22.7	25.8
.600	.....	1.02	20.2	23.2
0.700	.....	1.01	13.7	15.6

6000°, 6200°, and 6400° K have been drawn for comparison. These were obtained by use of Planck's radiation law, reduced to mean solar distance. The application of the coefficient e to the measures most nearly approximates 6200° for integrated light and 6400° for the center of the disk. In the ultraviolet, however, the corrected energy-curve of the sun still falls far below that of a black body.

Since the thermoelectric measures at λ 0.39 μ include the K line and those at λ 0.40 μ the H line, the effect of the absorption of the two together is smoothed out. To estimate the intensity at λ 0.395 μ, we take the quotient of the ordinates of the smoothed curves of

e Av. *i* and *e* Av. *c* at that wave length and *e* for λ 0.395 μ. These, plotted upon the original energy-curves, give the depressions at λ 0.395 μ seen in Figure 3. It is significant that the H and K lines seem to play little part in the great drop of intensity between λ 0.40 and λ 0.38 μ. There are no lines in the ultraviolet solar spectrum attributable to terrestrial origin until we reach the Huggins band of ozone superposed upon the great absorption band at the violet end of the solar spectrum. The Huggins band consists principally of a cluster of pairs of broad lines shading to the red, spaced about 28 A apart, with the individual lines of each pair separated about 7 A. The width of the lines themselves is about 8 A, and of a pair about 13 A, the shading of the violet component overlapping the red. Since their intensities are rather low, the lines are entirely masked by lines of solar origin, even with low dispersions.

In 1926 the author made a laboratory study of the ozone spectrum, using the first order of a 1-m concave-grating spectrograph, mounted Eagle fashion (dispersion 17.4 A/mm). The ozone tube, 67 cm long, closed by crystal quartz windows, was charged from a Berthelot generator operated from a 50,000-volt transformer. With this equipment an ozone density could be obtained equivalent to a layer 1.5 cm thick at N.T.P. The analysis of the ozone content of the tube was made by the iodine method. The exposures were on Cramer Contrast plates, with light from the crater of the carbon arc. An iron-arc comparison spectrum was added to each exposure by using a focal-plane occulting bar. Several exposures were impressed on each plate together with exposures through a calibration wedge for intensities.

From this material four spectra referring to a variety of ozone densities were selected as having the best exposures. These were measured direct and reverse and referred to the comparison lines in the usual way. The results are given in Table 5. The first column gives the wave lengths, those in parentheses being measured by Fowler and Strutt[12] only. Three lines not given by Fowler and Strutt were found at λλ 3148.6, 3213.9, and 3440.2. Since the accuracy of measurement cannot be very great, only the nearest 0.1 A is recorded. The second column gives the mean residuals for the

[12] *Proc. R. Soc.*, **93**, 577, 1917.

four spectra, and the third the differences between the present measures and those of Fowler and Strutt. The strong lines bracketed in the first column are members of pairs, and their positions are marked in Plate I by their intensity numbers. The braces in this plate show the region of ozone absorption for each pair of lines. It appears that the Huggins band only partly accounts for the general

TABLE 5

WAVE LENGTHS AND LINE INTENSITIES IN THE ULTRAVIOLET OZONE BAND

λ	Mean Residuals	Mt. W. *minus* F and S	*I*		λ	Mean Residuals	Mt. W. *minus* F and S	*I*	
(3089.5)	.....	.....	8		3242.0	0.1	−1.0	1	
(3096.5)	.....	.....	4		3248.5	.1	−1.2	6	(0.10)
(3105.0)	.....	.....	5		3255.0	.2	−0.5	6	
3114.7	.....	+0.4	8		3271.9	.1	−0.1	2	(0.09)
3137.2	0.6	−0.2	10		3279.6	.1	−0.2	5	
3148.6	.1	.....	0		(3284.0)	.....	.....	2	
3156.0	.2	−0.1	8		3303.8	.2	−0.3	3	(0.06)
3163.0	.2	+0.4	2		3311.0	.5	−0.5	4	
3171.6	.1	0.0	4	(0.18)	3330.8	.2	−0.4	1	(0.05)
3176.3	.1	−0.3	8		3337.4	.1	−1.1	3	
3181.1	.1	−0.4	1		(3346.0)	.....	.....	1	
3187.4	.....	−1.4	1		3363.2	.3	−1.0	1	(0.03)
3194.8	.2	0.0	7	(0.14)	3372.4	.1	−1.7	2	
3200.4	.1	+0.6	7		(3377.7)	.....	.....	1	
3206.4	.2	−0.4	0		3402.2	.7	−0.4	1	
3213.9	.2	.....	0		(3421.4)	.....	.....	1	
3220.6	.0	−0.9	10	(0.15)	(3432.2)	.....	.....	1	
3227.0	0.0	−0.2	9		3440.2	0.5	.....	0	
(3232.8)	.....	.....	1						

absorption at λλ 3193, 3213, 3227, 3250, and 3281 and that these depressions are mostly due to groups of the stronger solar lines.

The intensities in the fourth column are on a scale of 10. For wave lengths shorter than λ 3187 they are taken from Fowler and Strutt (save λ 3148.6, which is not in their list), who used greater dilutions of ozone. Values in parentheses, each referring to a pair of lines, were obtained by microphotometry, the calibration wedge being used to convert the photometric curves into intensities. The intensity ratio with ozone in the tube to that with oxygen was formed for each wave length. Plotted against wave length, they

formed a curve with depressions for each ozone line. The ratio of the depth of each depression to the ordinate of the smoothed curve is then the absorption of the line.

The sums of the estimated intensities for a pair of lines approximate the measured intensities for 1 cm of ozone, if we disregard the decimal. Since the maximum in absorption is 18 per cent and averages 10, we would expect only about 1 per cent absorption in the bands in the region λλ 0.32–0.33 μ indicated by the record in Plate I, which was made at air mass 1.3 on Mount Wilson. From this discussion we see that the Huggins band of lines will be scarcely distinguishable in the clusters of Fraunhofer lines in a G-type spectrum.

PRELIMINARY LINE-GROUP INTENSITIES

The individual lines are not sufficiently separated in the photoelectric energy-curves from the 21-foot monochromator (Pl. I) to permit accurate measures of their intensities. Such curves, however, furnish material for a study of the intensities of groups of lines included between wave lengths at which the continuous spectrum approaches the general contour of the energy-curve. The 100-A strips for λλ 3900–4000 and λλ 3200–3300 A were accordingly subdivided as shown in Plate I, the adopted wave-length limits being those given in Table 6.

Mulders[13] has defined the absorption of a line in terms of the width of an equivalent totally absorbing rectangular line in thousandths of an angstrom (*EW* in mA). We may determine *EW*, the line-group absorptions (intensities) according to this definition, by the formula

$$EW = 1000\,\Delta\lambda\,A_l\,A_c^{-1}, \qquad (12)$$

where $\Delta\lambda$ is in angstroms, A_l is the area of the line and A_c the area below the contour, both within the wave-length limits assigned. The results for *EW* from planimeter measurements are given in Table 6.

The corresponding line absorptions given under "Mulders' calibration" were obtained by summing for the lines within each interval the values obtained from Mulders' calibration-curves of the

[13] *Zs. f. Ap.*, **10**, 297, 1935.

Rowland intensities, which overlap the region λλ 3900–4000 A. Mulders' measured intensities were, however, used whenever given in his table.

TABLE 6

LINE-GROUP ABSORPTION IN TWO ULTRAVIOLET REGIONS

	Spectral Limits	*EW*	Mulders' Calibration	$\frac{P}{M}$	Spectral Limits	*EW*	Mulders' Calibration	$\frac{P}{M}$
		mA	mA			mA	mA	
	3894.5–97.0	775	656	1.39	3969.0–70.8	54	802	0.07*
	97.0–98.8	396	501	0.79	70.8–71.7	36	223	0.16
	98.8–01.2	816	908	0.90*	71.7–72.7	210	216	0.97
	3901.2–04.4	1056	1206	0.88*	72.7–75.4	459	826	0.56*
	04.4–07.4	1380	1520	0.91*	75.4–77.2	360	466	0.77
	07.4–09.5	504	533	0.95	77.2–79.0	288	382	0.76*
	09.5–11.6	525	644	0.81	79.0–80.0	100	227	0.44
	11.6–12.9	195	294	0.66	80.0–85.5	1045	1244	0.84
	12.9–15.2	460	687	0.67	85.5–88.0	550	570	0.96*
	15.2–18.2	690	832	0.83*	88.0–90.8	560	546	1.03
	18.0–20.0	513	615	0.84	90.8–93.6	392	505	0.78
	20.0–22.5	1363	968	1.41*	3993.6–00.4	1564	1516	1.03
	22.5–24.0	494	459	1.07	3189.5–98.9	4251	.....	.....
	24.0–25.0	156	205	0.76	98.9–04.8	1416	.....	.....
	25.0–26.7	294	522	0.56	3204.8–06.5	255	.....	.....
	26.7–29.0	630	582	1.08*	06.5–19.0	3750	.....	.....
	29.0–32.0	506	581	0.87*	19.0–21.0	320	.....	.....
K	15.0–53.0	11820	15900	0.75	21.0–24.5	980	.....	.....
	35.8–38.0	28	172	0.16	24.5–32.5	2880	.....	.....
	38.0–40.0	216	292	0.74	32.5–38.5	2160	.....	.....
	40.0–40.8	24	68	0.35	38.5–40.7	550	.....	.....
	40.8–42.8	360	629	0.57	40.7–46.5	928	.....	.....
	42.8–46.0	1024	1189	0.86*	46.5–55.5	2790	.....	.....
	46.0–49.5	600	955	0.63	55.5–59.5	440	.....	.....
	49.5–51.0	126	224	0.56	59.5–62.2	297	.....	.....
	51.0–54.9	720	978	0.74*	62.2–65.9	666	.....	.....
	54.9–57.5	672	810	0.83*	65.9–75.0	924	.....	.....
	57.5–59.1	288	315	0.92	75.0–89.9	3984	.....	.....
	59.1–64.9	1500	1617	0.93*	89.9–93.4	516	.....	.....
	64.9–67.0	1003	480	2.09	93.4–97.5	468	.....	.....
H	3953.0–82.0	10050	.....	.....	3297.5–00.0	510	.....	.....

* Region includes lines of Rowland intensity greater than 5 or more than one of intensity 5.

The column P/M gives the ratios for the two series of group intensities. The average ratio is 0.84, but the scatter is very large. If we examine the 15 ratios (indicated by asterisks) for which the spectral intervals include lines of Rowland intensities greater than 5,

or more than one line of intensity 5, the scatter is much reduced save for the interval λλ 3969.0–3970.8 A near the H line, where the photoelectric curves give no indication of intensities as high as 10 and 5 assigned by the Rowland table. The scatter is no doubt indigenous in the Rowland intensities themselves in this difficult region dominated by the H and K lines where intensities of other lines must be measured from the contours of the wings of H and K.

The intensity of the K line is 11,820 mA while H is 10,050 mA, a ratio of 1.18 against the theoretical ratio $\sqrt{2} = 1.41$ adopted by Mulders to obtain the value of H from that of K.

CARNEGIE INSTITUTION OF WASHINGTON
MOUNT WILSON OBSERVATORY
October 1939

GALACTIC STAR CLUSTERS[1]

ROBERT J. TRUMPLER

ABSTRACT

The principal methods for determining intragalactic distances greater than 500 parsecs are discussed and their application to galactic star clusters is outlined. A comparison of distance moduli derived from photometric spectroscopic data with distances based on angular diameters yields an average coefficient of photographic absorption of $0^m.7$ per kiloparsec. Preliminary results for the radial velocities of 39 clusters lead to a value $A = +0.015$ km/sec · parsec for the coefficient of differential galactic rotation. The good agreement of this result with other determinations confirms the correctness of the adopted distance scale of galactic star clusters. The radial velocities also indicate a negative k term increasing in amount with distance.

The appearance of galactic star clusters seen at great distances projected on the background of faint stars is studied. At distances of more than 5000 parsecs even the most favorable objects would have been missed by the surveys of the past, while the poorer and less luminous objects may have been overlooked at much smaller distances. This explains why the space distribution of the known star clusters shows little relation to the structure of our galaxy. Attention is drawn to the promise of a search for more distant clusters on photographs taken in red light with instruments of large aperture and large field.

I. INTRODUCTION

Among the many problems with which the study of galactic star clusters is concerned it seems appropriate here to take up those which bear directly on the structure and dimensions of our galactic star system. On the one hand, we have the distance determination which involves the study of interstellar light-absorption and galactic rotation, and, on the other hand, the space distribution of known clusters and the question of incompleteness in our present knowledge of these objects.

For the determination of galactic distances greater than 500 parsecs the following methods are available:

I. Photometric methods
   a) Luminosity from spectral class
   b) Luminosity from periods of variable stars

   (Requires a knowledge of the relation between absolute magnitude and spectral class, or the period-luminosity relation, and of the absorption of light in interstellar space.)

[1] This paper was presented at the Symposium on Galactic and Extragalactic Structure, held in connection with the dedication of the McDonald Observatory on May 5–8, 1939.

II. Angular diameters of extended objects
  a) Galactic star clusters
  b) Planetary nebulae
    (Requires a knowledge of the linear diameters of various classes of such objects.)
III. Differential galactic rotation effect on radial velocities
    (Requires a knowledge of the Oort constant A.)

Strictly speaking, all these methods are of a statistical nature; the first two, however, may within reasonable error limits be applied to individual objects, while the third can be used only to find the mean distance for groups of galactic objects. Moreover, these methods are not absolute; they require for calibration a knowledge of certain constants or statistical relations which have to be established empirically. Some of these, like the luminosity-spectral class relation or the zero point of the period-luminosity relation, result from a study of the nearer stars for which trigonometric and proper-motion parallaxes are available. Others, such as the average coefficient of interstellar light absorption, the linear diameters of clusters and nebulae, or the galactic rotation constant cannot be reliably determined from the nearer objects and have to be studied by intercomparing the results of different methods of distance determination. The distances of the more remote galactic objects depend considerably on the values adopted for the absorption coefficient and for the galactic rotation constant A. Star clusters offer the possibility of applying all three methods—a particularly favorable opportunity for the calibration of the distance scale.

II. THE PHOTOMETRIC-SPECTROSCOPIC DISTANCE DETERMINATION OF STAR CLUSTERS

The procedure for the photometric-spectroscopic distance determination of a galactic star cluster is as follows:

1. The apparent photographic magnitudes m and the spectral classes s of the cluster stars are found by direct observation.

2. After the giant stars have been separated from those of the main sequence and after the background stars have been eliminated as far as possible, the mean apparent magnitude $\bar{m}_s$ of the cluster stars is formed for each spectral class. This is compared with the

mean absolute magnitude $\overline{M}_s$ of the same spectral division in the Hertzsprung-Russell diagram of the nearer stars.

3. The distance modulus y of the cluster is given by

$$y = \overline{m}_s - \overline{M}_s ,$$

and its mean value for all cluster stars is adopted.

4. The relation between the distance modulus y and the distance r (in parsecs) is

$$y = 5 \log r - 5 + A(r) , \qquad (1)$$

where $A(r)$ is the amount of interstellar absorption affecting the light of a star at distance r.

The determination of the distance modulus involves the assumption that the luminosity-spectral class relation in a cluster is the same as for stars in general. Kuiper[2] suggests that the chemical composition (hydrogen content) of different members in the same cluster is very similar, while it may vary considerably from one cluster to another. A difference in hydrogen content, however, affects both the spectral class and the luminosity in such a way that a star of the main sequence is principally displaced parallel to the sequence. Although our assumption may not be strictly fulfilled for individual clusters, the error introduced will, in general, be small.

The absorption law $A(r)$, which must be known in order to find the distance r from the modulus y, does in reality vary with galactic longitude and latitude. Our information on the irregularities of the interstellar medium is as yet too incomplete to allow for these. Two simplifying assumptions about the absorption may be introduced as an approximation:

1. For the true absorption correction we substitute its statistical average for low galactic latitudes, and this average absorption is assumed to be uniform,

$$A(r) = \alpha \cdot r , \qquad (2a)$$

where α is a constant representing the average absorption of photographic light per kiloparsec near the galactic plane.

2. The photographic absorption $A(r)$ may be taken as proportional to the color excess E, which is the difference between the photo-

[2] *Ap. J.*, **86**, 176, 1937.

graphic and the visual absorption and can be ascertained by observation.

$$A(r) = \chi \cdot E\,, \qquad (2b)$$

where the constant χ is the ratio of photographic absorption to color absorption.

The second assumption is probably a much better approximation than the first, since it takes some account of the irregularities in the absorbing medium. The data on color indices in star clusters are, however, at present not sufficiently complete or sufficiently homogeneous for its application. In either case the constants α or χ must be evaluated empirically with the aid of one of the other methods of distance determination which is not affected by interstellar absorption.

III. DETERMINATION OF DISTANCES AND OF THE ABSORPTION COEFFICIENT FROM ANGULAR DIAMETERS

The diameter method is based on the principle that star clusters which have the same constitution C have the same linear diameter D_c (in parsecs), whatever their distance from the observer. Clusters of the same constitution can be grouped together by a classification which takes into account the number of cluster stars as well as the degree of central concentration. If the angular diameter d (in minutes of arc) of a cluster has been estimated by the observer, the distance r is obtained by

$$r = \frac{D_c}{d}\, 3438$$

or

$$\log r = \log D_c - \log d + 3.536\,. \qquad (3)$$

For each class of clusters the value of D_c has to be found by means of clusters with known distances. The photometric distances of the nearer clusters for which the absorption is negligible might be used for the purpose; but the number of such objects is rather too small. If the assumption in equation (2a) of uniform absorption is adopted, the constant α and the values of D_c can be determined by a combination of the photometric method and the diameter method.

Each cluster furnishes two equations of the forms (1) and (3):

$$\left.\begin{aligned} 5 \log r + a \cdot r &= y + 5 \\ -\log r + \log D_c &= \log d - 3.536 , \end{aligned}\right\} \tag{4}$$

in which the terms on the right-hand side are given by observation. For n clusters divided into k classes there are thus $2n$ equations with n unknown distances r; k unknown D_c's; one unknown a. Since k is much smaller than n, there are fewer unknowns $(n + k + 1)$ than there are equations of condition, and a statistical solution is possible. Owing to the transcendent nature of the first equation, however, the solution has to be accomplished by successive approximations.

The first step is to find the relative values of the mean linear diameters

$$D'_c = \frac{D_c}{D_0} \tag{5}$$

of the different classes, where the constant D_0 represents an unknown scale correction to the adopted relative diameters D'_c.[3] When the value of $\log r$ from equation (3) and of D_c from equation (5) are introduced into equation (1), the latter becomes

$$A(r) + c = y + 5 \log \frac{d}{D'_c} - 12.68 , \tag{6}$$

where

$$c = 5 \log D_0 .$$

It is thus possible to compute for each individual cluster the amount of photographic absorption, except for a constant zero correction on which the scale of the distances by the diameter method depends.

For 97 clusters[4] for which reliable and homogeneous data are available, the values of $A(r) + c$ have been plotted in Figure 1 as a function of the distance modulus y, which is also a function of r and is obtained by the photometric method. The vertical scattering of

[3] In the present solution these were taken from *Lick Obs. Bull.*, **14**, 160, 1930.

[4] Of the 100 clusters listed in *ibid.*, p. 156, Table 3, many of the southern clusters with incomplete data were omitted, while clusters with more recent observations were added; about one-third of the clusters are new or have new data.

the points is, in the first place, due to observational errors in the estimated angular diameters and in the classification which enters into D'_c. As these are approximately percentage errors, the results for $A + c$ are practically of the same accuracy for near and distant clusters. The errors in y, as far as they are due to incompleteness of data or statistical uncertainty of the photometric method, are considerably smaller. Part of the scattering of points is probably due also to irregularities in the absorption. Normal points found for

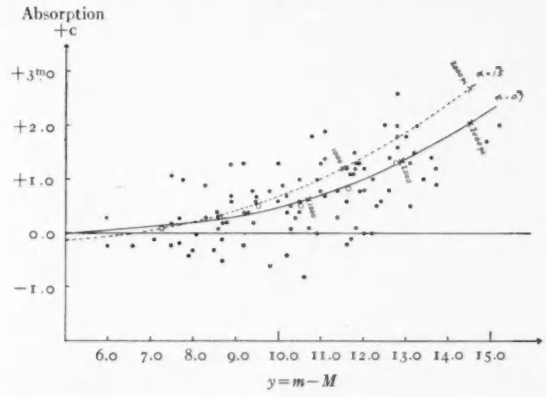


FIG. 1.—Photographic absorption of star clusters as function of distance modulus. The abscissa represents the distance modulus obtained by the photometric method, the ordinate the absorption (plus an unknown zero correction c) derived from equation (6). Each cluster is marked by a full dot, group means by open circles. The full curve shows the relation between absorption and distance modulus in case of a uniform absorption with a coefficient $\alpha = 0^m.70$ per kiloparsec, the dotted curve for a coefficient $\alpha = 1^m.5$ per kiloparsec. The unknown zero correction c is for both curves so determined as to fit the nearer clusters.

groups of 12–20 clusters are marked by open circles; the full-line curve shows the relation between $A + c$ and y for uniform absorption and a coefficient

$$\alpha = 0^m.70 \text{ per kiloparsec},$$

which gives the best fit with the observations. The dotted curve indicates the case of uniform absorption with a coefficient $\alpha = 1^m.5$ per kiloparsec; for distances greater than 700 parsecs such a large coefficient seems quite incompatible with our data.

In a similar way it would be possible with the assumption of equation (2*b*) to determine the unknown ratio χ and the values of D_c by a combination of the photometric method and the diameter method.

IV. DISCUSSION OF SYSTEMATIC ERRORS

To obtain a correct result for the absorption coefficient by this method it is essential that the estimates of angular diameters and

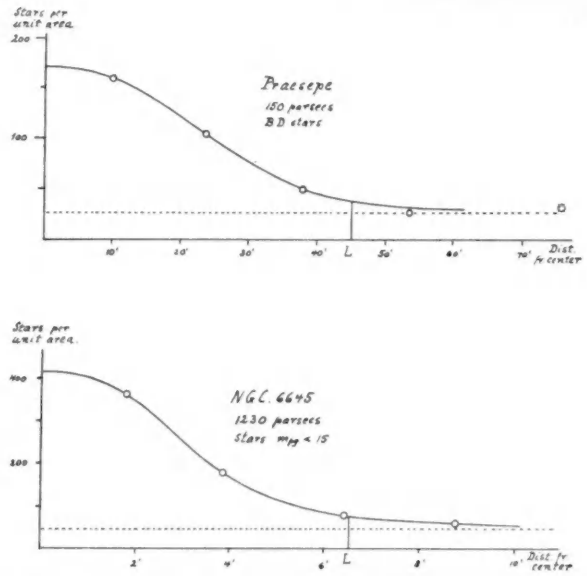


FIG. 2.—Areal star density in Praesepe and NGC 6645. The number of stars per unit area (ordinate) is plotted as a function of the angular distance (abscissa) from the center of the cluster. The dotted horizontal line indicates the mean density of the background stars; *L*, the estimated limit of the cluster.

classification be homogeneous and free from systematic errors depending on the distance of the objects.

In estimating the angular diameter, we have attempted to find the diameter of the smallest circle outside of which no excess of cluster stars over the background stars is noticeable by mere examination of a photograph or chart. It is of interest to compare the results of such estimates with those of star counts. Figure 2 gives such a comparison for Praesepe and NGC 6645, two clusters of similar constitution (class 1*r*) but at very different distances. From star counts in con-

centric rings drawn around the center of each cluster the star density per unit area was derived and plotted against the angular distance from the center. For Praesepe the counts were made on the *Bonner Durchmusterung* chart, for NGC 6645 on a photograph reaching the photographic magnitude limit 15. The "estimated" limit of the cluster is indicated by the vertical line at L and is considerably inside of the true limit shown by the star counts. The scale of the abscissas is so chosen that the estimated limiting radius has the same length for both clusters, and the circle with this radius is taken as unit area. The scale of ordinates is made inversely proportional to the total number of cluster stars within the limiting circle, and the areal density of the background stars is given by the dotted horizontal line. The two characteristic curves obtained are remarkably similar for the two clusters.

This and other tests show that if the observer enlarges different clusters of the same class to approximately the same size and increases the magnitude limit of the photographs accordingly, the estimated angular diameters are well comparable and are little affected by the distance of the cluster. Since the conditions stated have, on the whole, been realized, a direct effect of distance on the estimates of diameter and central concentration such as suggested by Ten Bruggencate[5] is out of the question.

It will be noticed in Figure 2 that the observer's estimate overlooks that part of the cluster where the excess of cluster stars becomes less than 50 per cent of the background stars. The richness of the background on which a cluster appears projected may, therefore, have an effect on the diameter estimates. Local irregularities in the star field of the Milky Way affect the statistical data like accidental errors of observation. But the question we have to investigate is whether the number of background stars within the limiting circle increases with increasing distance of the observer when the magnitude limit is raised according to the distance.

The example of Praesepe will illustrate this point. On a photograph with the apparent photographic magnitude limit 9.6, Praesepe will appear as a group of 35 stars occupying a circle 90′ in diameter, and such a field will contain 9 additional background stars. If we

[5] *Naturwissenschaft.*, **18**, 725, 1930.

imagine this cluster removed ten times farther away (distance 1500 parsecs) along the galactic plane, it will reduce to an area of 9′ in diameter, and to show the same 35 cluster stars our photograph must reach the magnitude limit 15.7 (assuming an absorption of $0^m.8$ per 1000 parsecs). In low galactic latitudes the average number of stars brighter than $15^m.7$ in a circle of 9′ diameter is 30, and the 35 cluster stars will now appear projected on a field containing 30 background stars.

If the same upper limit of absolute magnitude is reached in all clusters, we find in low galactic latitudes the more distant clusters projected on a richer background. The angular diameters of distant clusters are thus probably estimated somewhat too small as compared with the nearer clusters. On the other hand, a richer background would also lead an observer to underestimate the number of cluster stars and would affect the classification. As the adopted linear diameters D_c are smaller for the poor clusters than for the rich clusters, an underestimate of the angular diameter d for distant clusters would in many cases be accompanied by the assignment of too small a linear diameter D_c. The ratio $D_c : d$ which determines the distance by the diameter method will thus, on the whole, be much less affected. If the two effects do not quite compensate each other, it is more likely that the diameter distances of the distant clusters are slightly too large and the absorption coefficient a derived therefrom slightly too small. We have therefore retained the value $a = 0^m.8$ per kiloparsec, with which the distances of *Lick Observatory Bulletin*, No. 420, were computed.

Van Rhijn[6] suggested that the clusters with spectroscopic data suitable for such an investigation are necessarily selected so as to favor objects seen through the more transparent regions of our stellar system. This is undoubtedly true for clusters more distant than 1500 parsecs, and the average photographic absorption coefficient in low galactic latitudes may well be as large as $1^m.0$ per kiloparsec. Even if such a selection is admitted, the absorption coefficient determined from the clusters is that which applies to these objects on the average and must be used for the correction of the photometric distance measures.

[6] *Groningen Pub.*, No. 47, 1936.

V. GALACTIC ROTATION SOLUTION

The third method, based on the differential galactic rotation effect in radial velocity, can now be used to compare our distance scale of star clusters with that of other galactic objects (B stars, Cepheid variables, planetary nebulae). Although the radial velocity measures in 75 galactic star clusters are far from complete, preliminary data are at present available for 39 clusters with distances ranging from 430 to 4600 parsecs.

The observed radial velocities of the clusters, corrected for solar motion (19.5 km/sec toward $A = 18^h.0$, $D = +30°$), which we shall designate as V, were represented by the formula

$$V = k_1 + k_2 [r \cos^2 b] + A[r \cos^2 b \sin 2(l - 325)] ,$$

in which r is the adopted distance (mean of photometric and diameter methods), l and b are the galactic longitude and latitude of each cluster, while the longitude of the galactic center was assumed to be $325°$. The last term represents the effect of differential galactic rotation when higher harmonics are neglected. The k term was assumed to consist of a constant k_1 and a term proportional to $r \cos^2 b$. The determination of three unknown constants by a least-squares solution furnished the following values:

$$A = +0.0150 \pm 0.0010 \text{ km/sec} \cdot \text{parsec} ,$$

$$k_1 = -3.5 \text{ km/sec} ,$$

$$k_2 = -0.0043 \text{ km/sec} \cdot \text{parsec} .$$

The residuals of the observations from the formula give a dispersion (standard deviation) of ± 12 km/sec for the peculiar velocities of galactic star clusters. That the value of k_1 mostly represents a systematic error in our preliminary radial velocities is confirmed by observations of bright standard stars. The term with k_2, however, appears to be real and suggests a local contraction along the galactic plane of the observed star clusters. A similar change of the k term with increasing distance is indicated by the radial velocities of O and B stars,[7] and it is remarkable that the galactic rotation solutions of

[7] See Paris Pismis, *Pub. Istanbul U. Obs.*, No. 10, 1938.

other distant galactic objects, such as planetary nebulae[8] and Cepheid variables,[9] also lead to negative k terms.

Table 1 lists the results obtained by various observers for the galactic rotation constant A (third col.) and for the average coefficient of photographic absorption α per kiloparsec (fourth col.). The second column indicates the class of objects studied, and the last column, the methods used for their distance determinations. Our result for A is in close agreement with that of Plaskett and Pearce[10] and with that

TABLE 1

GALACTIC ROTATION CONSTANT AND MEAN ABSORPTION COEFFICIENT

Author	Objects	A (km/sec · parsec)	α (per kpc)	Distance Scale
Plaskett, Pearce*	O-B3 stars	0.0155±0.0009	.....	Proper-motion parallaxes
Oort†.....	B, c stars	.019 ±0.003	1^m8	Proper-motion parallaxes
Joy‡.....	Cepheid variables	.0209±0.0008	0.85	Period-luminosity relation
Berman§.....	Planetary nebulae	.0140	0.72	Proper-motions; photometric; angular diameters
Trumpler.....	Galactic star clusters	0.0150±0.001	0.70	Photometric; angular diameters

* *Pub. Dom. Ap. Obs.*, **5**, No. 3, 1933.
† For A: *B.A.N.*, **4**, 82, 1927; for α: *ibid.*, **8**, 233, 1938.
‡ *Mt. W. Contr.*, No. 607, 1939.
§ *Lick Obs. Bull.*, **18**, 64, 1937.

of Berman but differs considerably from that of the Cepheid variables.

A former galactic rotation solution by Miss Hayford,[11] based on 25 galactic clusters, yielded a considerably smaller value for A (0.009 km/sec · parsec), and this has induced some investigators[12] to suggest a scale error in our cluster distances. The difference between the result of the new solution and the former one is partly due to the additional and improved data and partly to the different treatment. Miss Hayford's solution included a constant k term as well as the elements of the solar motion and led to a large value of 30 km/sec for the solar velocity. Our new solution adopts the usual value of the

[8] *Lick Obs. Bull.*, **18**, 64, 1937.
[9] *Mt. W. Contr.*, No. 607, 1939.
[10] *Pub. Dom. Ap. Obs.*, **5**, No. 3, 1933.
[11] *Lick Obs. Bull.*, **16**, 53, 1932.
[12] Raimond, *Groningen Pub.*, No. 46, p. 32, 1934; Oort, *B.A.N.*, **8**, 245, 1938.

solar motion but introduces a k_2 term proportional with the distance in addition to a constant k_1.

VI. THE OBSERVED SPACE DISTRIBUTION OF THE CLUSTERS

The distances of all known galactic star clusters (334) which are sufficiently well defined were derived by either or both of the methods outlined in sections II–III, and these data permit a study of the

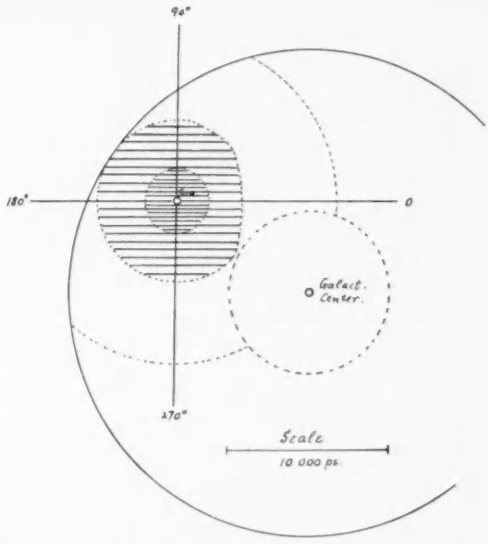


FIG. 3.—Distribution of known galactic star clusters in the projection on the galactic plane. Within the circle of dark shading the clusters are nearly uniformly distributed; in the ring of light shading they thin out rapidly. The large dotted circle drawn around the sun indicates the limit of 10,000 parsecs to which a search of clusters might be extended; the full circle shows the assumed limit of the galactic system.

space distribution. Figure 3 gives an outline of the region (shaded area) occupied by the known clusters as compared with the outline (large circle) of the galactic system according to the generally accepted hypothesis. The projection of the individual clusters on the galactic plane is illustrated in *Lick Observatory Bulletin*, No. 420, p. 179. Essentially, the clusters group themselves around a center situated near the sun, and their arrangement shows no relation to the general structure of the galaxy. Up to a distance of 2000 parsecs from the sun the clusters are in their projection on the galactic plane

fairly uniformly distributed. As the distance increases beyond 2000 parsecs they thin out rapidly, and at a distance of 5000 parsecs we practically reach the limit of the known clusters.

The question immediately imposes itself as to whether these features in the apparent distribution mainly reflect the incompleteness of our present cluster list. To admit such an interpretation it must be shown, in the first place, that even the richest and most luminous clusters would not be observable at distances greater than 5000 parsecs with the instruments used for the search, such as Herschel's reflector or the Franklin-Adams camera. In the second place, an explanation has to be given of how a large number of clusters could have escaped discovery even at distances of 2000–3000 parsecs.

The well-known local clouds of obscuration (dark nebulae) in the Milky Way are undoubtedly obstructing the view of some of the more distant clusters, but their effect can hardly explain the rapid thinning-out, starting at 2000 parsecs, or the lack of clusters beyond 5000 parsecs. The existence of a general absorption of light in a layer stretching along the galactic plane will, on the other hand, greatly affect all distant clusters. Although this layer appears to be of considerable irregularity, we have assumed an average rate of $0^m.8$ per 1000 parsecs for the absorption of photographic light.

To study the extreme distance limit to which galactic clusters are observable on Milky Way photographs with a given equipment, we have selected three of the richest and most luminous clusters: Praesepe, the Pleiades, and h Persei. The data of Table 2 illustrate how these clusters would appear on a photograph containing stars brighter than magnitude 18, if seen from distances of 3000, 5000, and 7500 parsecs. The third column of the table shows how much fainter the cluster stars would be than at their actual distances. The fourth column gives the apparent angular diameter, column 5 the number of cluster stars brighter than eighteenth magnitude, and column 6 the number of foreground or background stars within the field of the cluster. Column 7 indicates the percentage of all stars within the area of the cluster which are cluster members; a cluster would probably not be recognized as such if this percentage is less than 25 or 30. The magnitude of the brightest cluster star is found in the last column.

Praesepe would evidently no more appear as a cluster at a dis-

tance of 5000 parsecs; the Pleiades would still be recognized at 5000 parsecs but not beyond; h Persei might possibly be noticed at a distance of 6000 parsecs. None of the systematic surveys of the Milky Way, such as the Franklin-Adams chart, reaches the eighteenth-magnitude limit, and these most favorable examples leave little doubt that clusters more distant than 5000 parsecs would have been discovered in only a few exceptional cases.

We have now to examine the effect of distance on some of the poorer clusters containing no stars of high luminosity. The star clus-

TABLE 2

APPEARANCE OF CLUSTERS AT GREAT DISTANCES

(Absorption $0^m.8$ per 1000 Parsecs)

Cluster	Distance (Parsecs)	Δm (Pg)	Estimated Angular Diameter	$m_{pg} < 18$		Per Cent Cluster Stars	Brightest Cluster Star
				Cluster Stars	Background Stars		
Praesepe	3000	8.8	4'.5	26	44	37	15.0
	5000	11.5	2.7	2	16	.....	17.7
Pleiades	3000	8.8	6	49	78	39	11.7
	5000	11.5	3.6	15	28	35	14.4
	7500	14.4	2.4	1	12	.....	17.3
h Persei	5000	5.9	5.2	97	59	62	12.7
	7500	8.8	3.5	5	27	16	15.6

ter in Coma Berenices is the best-known example of this type. Table 3, in an arrangement similar to Table 2, describes the appearance of the cluster as seen at various distances, projected on an average star field in low galactic latitude. Three photographs with limiting magnitudes of 18, 16, and 14 are assumed to be available. Although the five brightest cluster stars would be visible on the first photograph at a distance of 4000 parsecs, they would be lost among the great number of background stars. On the second photograph the percentage of cluster stars is somewhat more favorable for distances of 2000–3000 parsecs, and on the third photograph the cluster would be fairly noticeable up to a distance of 1000 parsecs. On a long exposure a poor cluster at great distance is "drowned" by the background and foreground stars; on a shorter exposure the cluster stars

are only visible if the distance is not too great. From the data of Table 3 we conclude that the Coma Berenices cluster might be recognized as a clustering to a distance of 1500 or 2000 parsecs. For poor clusters without stars of high luminosity the distance limit to which they are noticeable is mainly conditioned by the rapid increase of background stars with increasing magnitude limit, and most of such clusters would be overlooked at distances greater than 2000 parsecs.

The observed space distribution of galactic clusters seems, therefore, satisfactorily explained by incompleteness and selection of the known objects.

The absorption of light in the vicinity of the galactic plane is the most serious obstacle in any search for more distant galactic clusters.

TABLE 3

APPEARANCE OF COMA BERENICES CLUSTER AT DIFFERENT DISTANCES

Distance (Parsecs)	Δm (Pg)	Angular Diameter	Brightest Cluster Star	$m_{pg} < 18$			$m_{pg} < 16$			$m_{pg} < 14$		
				Cluster Stars	Background Stars	Per Cent	Cluster Stars	Background Stars	Per Cent	Cluster Stars	Background Stars	Per Cent
1000.....	6.4	18′	11.6	31	700	4	23	147	14	12	26	32
2000.....	8.7	9	13.9	21	180	10	11	37	23	3	6	.....
3000.....	10.4	6	15.6	12	78	13	5	16	24	.....	.....	.....
4000.....	11.8	4.5	17.0	5	44	10	.....	.....	.....	.....	.....	.....

As this absorption is greater for photographic light than for longer wave lengths, it is of interest to investigate whether observation in yellow or red light would be more advantageous for the discovery of distant galactic objects. That this is the case is clearly indicated by the data of Table 4, which describes the appearance of Praesepe, the Pleiades, and h Persei at distances of 5000–10,000 parsecs, as seen on photographs taken in red light (λλ 6000–6500), reaching the seventeenth red magnitude. On such photographs, clusters like the Pleiades may be recognized at a distance of 7500 parsecs, and such as h Persei even somewhat farther. The red magnitude limit 17 is thus more effective than the photographic magnitude limit 18, because the highly luminous cluster stars at such great distances are much reddened by selective absorption, much more than the background stars, which are in reality mostly foreground stars.

It is a remarkable fact that the introduction of photographic observing methods during the last fifty years has added very few new

galactic clusters to the large list of those discovered by the Herschels. Their use of visual observations evidently gave them an advantage over ordinary photographs utilizing light of short wave lengths.

An extension of our knowledge of clusters to a distance of 10,000 parsecs seems, at present, quite feasible if the Milky Way belt is photographed in red light with an instrument of sufficient resolution, light-gathering power, and a large field. A large Schmidt reflector would seem particularly adapted to such purpose and may possibly record stars as faint as the eighteenth or nineteenth red magnitude.

TABLE 4

APPEARANCE OF CLUSTERS AT GREAT DISTANCES IN RED LIGHT

(Absorption $0^m.45$ per 1000 Parsecs)

CLUSTER	DISTANCE (PARSECS)	ESTIMATED ANGULAR DIAMETER	$m_{red} < 17$		PER CENT CLUSTER STARS	BRIGHTEST CLUSTER STAR
			Cluster Stars	Background Stars		
Praesepe	5000	2'.7	12	16	43	15.6
Pleiades	7500	2.4	7	12	37	14.9
h Persei	7500	3.5	26	28	48	13.8
	10000	2.6	4	15	21	15.5

The diagram of Figure 3 gives a good idea of the results that might be gained by a systematic search for distant galactic clusters. A study of their space distribution to a distance of 10,000 parsecs from the sun (dotted circle) would bring out some of the significant structural features of our galaxy. In galactic longitudes 80°–200° the limit of the system could be traced, and in this direction we would not expect an appreciable number of more distant clusters. A point of great interest would be to ascertain whether the region of the galactic center is devoid of clusters. Such a possibility is suggested by an examination of the nearer spiral nebulae of intermediate and later types, in which nuclei of condensation or clusterings are mainly found in the spiral arms but not in or near the central nucleus. If our present views on the dimensions of our galactic system are correct, we must expect a great number of undiscovered clusters in galactic longitudes 0°–70° and 220°–290°.

UNIVERSITY OF CALIFORNIA

November 3, 1939

THE CONTINUOUS ABSORPTION OF LIGHT BY NEGATIVE HYDROGEN IONS

H. S. W. MASSEY AND D. R. BATES

ABSTRACT

In view of the possible importance of negative hydrogen ions in absorbing long-wave-length radiation in stellar atmospheres, it is desirable that a reliable theory of the continuous absorption by such ions be available. Such a theory is provided in this paper, and the resulting calculated values of the absorption coefficient in the wave-length region of astrophysical importance are shown to be accurate. This is made possible, on the one hand, by the existence of the accurate wave function for the ground state of H^- due to Hylleraas and, on the other, to a detailed analysis of the factors determining the form of the continuous wave function representing the state of the system in which one electron is ejected. The effects of distortion of the wave representing this electron by the static and polarized atomic field, of electron exchange, and of the non-separability of the wave function are all considered in detail and are found, for the particular case of H^-, to be incapable of affecting the calculated values.

It has been pointed out by Wildt[1] that, in discussing the absorption of radiation in the outer layers of stellar atmospheres, account must be taken of the possible existence of negative ions in such regions. In particular, in an atmosphere including both metal and hydrogen atoms the ionization of the former can supply electrons necessary for the formation of negative hydrogen ions. The energy of attachment of an electron to a hydrogen atom is 0.70 e. volts[2] (there is only one stationary state of the attached electron). As a result the presence of H^- ions provides a source of continuous absorption of long-wave-length radiation (wave No. 5674 cm^{-1} and upward) and Wildt[3] and Strömgren[4] have shown that this suggests a way out of certain difficulties concerning the magnitude and the wave-length dependence of stellar absorption coefficients. An accurate theory of the continuous absorption of light by H^- ions is therefore of considerable interest, and it is fortunately possible to develop such a theory because of the relative simplicity of the problem, which involves the interaction of two electrons only. Calculations in this direction have been already made by Jen[5] and by Massey

[1] *Ap. J.*, **89**, 295, 1939.

[2] Hylleraas, *Zs. f. Phys.*, **60**, 624, 1930.

[3] *Ap. J.*, **90**, 611, 1939.

[4] Unpublished.

[5] *Phys. Rev.*, **43**, 540, 1933.

and Smith,[6] but neither result can be regarded as indicating more than the order of magnitude of the effective cross-sections involved until a more detailed analysis is employed. Jen, although using an accurate wave function for the negative ion, did not use the correct form for the continuous wave function of the detached electron, while Massey and Smith, although using the correct form of continuous wave function, used only a rough approximation for that of the ion. In this paper the problem is treated accurately as one involving two electrons, and a number of effects not previously taken into account in the calculation of absorption coefficients are discussed in detail and are found to lead to no appreciable modification of the results.

GENERAL FORMULATION OF THE PROBLEM

Consider a stream of light quanta of definite frequency traversing a block of matter containing N atoms/cm^3. The fraction of quanta absorbed in a distance δx is given by $NQ\delta x$, where Q is the effective area or absorption cross-section for absorption of a quantum by an atom. The intensity of the beam in traversing a distance x is therefore diminished by a factor e^{-NQx}, and the absorption cross-section Q is identical with the atomic absorption coefficient of the material. Our problem is to determine Q for H^- ions as a function of the frequency ν of the incident quanta. This involves determining the probability of a photoelectric transition from the normal state of the ion to one of its continuous states. If E_0 is the electron affinity of the ion (the energy excess of the neutral atom over that of the negative ion), and E the energy $\frac{1}{2}mv^2$ of the ejected electron, we must have

$$E = h\nu - E_0 .$$

If $\Phi(\mathbf{r}_1, \mathbf{r}_2)$ is the wave function of the ground state of the negative ion and $\Psi_E(\mathbf{r}_1, \mathbf{r}_2)$ that for the state of energy $E + E_H$ (where E_H is the energy of the neutral hydrogen atom in its ground state), the absorption cross-section is given by the usual quantal formula

$$Q = \frac{32\pi^4 m^2 e^2}{3h^3 c} \nu v \, |\int\int \Phi^*(\mathbf{r}_1, \mathbf{r}_2)(\mathbf{r}_1 + \mathbf{r}_2)\Psi_E(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2|^2 , \qquad (1)$$

[6] *Proc. R. Soc.*, Ser. A, **155**, 472, 1936.

provided Φ and Ψ_E are suitably normalized, i.e., Φ to unity and Ψ_E to represent an electron bound to the proton, together with an outgoing current of unit density. This formula can certainly be relied upon to give accurate results provided the wave functions Φ, Ψ_E are accurately known. For most absorption problems the accurate determination of wave functions is a matter of some difficulty, but in the particular case of H^- we can obtain these in a satisfactory manner by methods which we now proceed to discuss.

DETERMINATION OF WAVE FUNCTIONS: THE GROUND-STATE WAVE FUNCTION OF H^-

A thorough study of the structure of H^- has been made by Hylleraas,[7] using the variation method which has been so successful in determining the energy in the two-electron problem of the helium atom. The wave function[8] which he gives has the form

$$\Phi(r_1, r_2) = Ne^{-\alpha(r_1+r_2)}\{1 + \gamma(r_1 - r_2)^2 + \beta r_{12}\}, \qquad (2)$$

where α, β, γ have the values 0.770, 0.308, 0.119, respectively. The term N is the normalizing factor given by

$$\alpha^3\left\{\pi\left(1 + \frac{35\beta}{8\alpha} + \frac{3\gamma + 6\beta^2}{\alpha^2} + \frac{77\beta\gamma}{8\alpha^3} + \frac{9\gamma^2}{\alpha^4}\right)r\right\}^{-(1/2)}$$

and

$$r_{12} = |r_1 - r_2| .$$

The corresponding wave function for the helium atom leads to a total binding energy of the two electrons equal to 78.45 e. volts, differing only very slightly from the observed 78.48 e. volts.[9] It is very unlikely, therefore, that the form in equation (2) is seriously in error in those regions of r_1, r_2 which contribute appreciably to the energy of the negative ion, i.e., in the region where $\Phi(r_1, r_2)$ is large. One cannot be so sure of its accuracy in other regions (particularly, at large distances from the nucleus), but it so happens that in the frequency range of astrophysical importance, the main contribution to

[7] *Loc. cit.*

[8] Throughout this paper the wave functions are given in atomic units.

[9] Baber and Hassé, *Proc. Cambridge Phil. Soc.*, **33**, 253, 1937.

the integral in equation (1) comes from such values of r_1, r_2 that we can be confident that the function is satisfactory for our purpose.

THE CONTINUOUS WAVE FUNCTION

In the absence of any interaction between the hydrogen atom and the ejected electron, we would have

$$\Psi_E(\boldsymbol{r}_1, \boldsymbol{r}_2) = 2^{-(1/2)}\{\chi_{10}(\boldsymbol{r}_1)e^{ikz_2} + \chi_{10}(\boldsymbol{r}_2)e^{ikz_1}\}, \tag{3}$$

where χ_{10} is the ground-state wave function of the hydrogen atom, $\pi^{-(1/2)}e^{-r_1}$, and k is the momentum of the ejected electron. The symmetrical wave function alone concerns us since the ground state of H^-, being a singlet, is also symmetrical in the space co-ordinates of the two electrons. This simple approximation can be shown to give results of quite high accuracy for the problem with which we are concerned, so we will consider first the calculation of the absorption coefficient with its use and then show that the corrections introduced by higher approximations to Ψ_E are unimportant.

CALCULATION OF ABSORPTION COEFFICIENT

We have, using equations (2) and (3),

$$Q = \frac{32\pi^4 m^2 e^2}{3h^3 c}\nu v|I|^2,$$

where

$$I = (2\pi)^{-(1/2)}N\iint e^{-a(r_1+r_2)}\{1 + \gamma(\boldsymbol{r}_1 - \boldsymbol{r}_2)^2 + \beta r_{12}\}(\boldsymbol{r}_1 + \boldsymbol{r}_2)(e^{ikz_2-r_1} + e^{ikz_2-r_1})d\boldsymbol{r}_1 d\boldsymbol{r}_2 .$$

Using the expansion

$$e^{ikz} = \left(\frac{\pi}{2kr}\right)^{1/2}\sum_n i^n(2n+1)J_{n+1/2}(kr)P_n(\cos\theta),$$

and carrying out the angular integrations involved, we can readily see that

$$I = (I_A + \beta I_B)16\pi^2 N k^{-(1/2)},$$

where

$$I_A = \int_0^\infty \int_0^\infty e^{-ar_1-(a+1)r_2}\{1 + \gamma(r_1 - r_2)^2\} r_1^{5/2} r_2^2 J_{3/2}(kr_1) dr_1 dr_2 ,$$

$$I_B = \int_0^\infty \int_0^\infty e^{-a(r_1+r_2)} r_1^3 r_2^2 \{\rho_0(r_1, r_2) e^{-r_2} r_1^{-(1/2)} J_{3/2}(kr_1) + \tfrac{1}{3}\rho_1(r_1, r_2) e^{-r_1} r_2^{-(1/2)} J_{3/2}(kr_2)\} dr_1 dr_2 ,$$

with

$$\left.\begin{aligned} \rho_0(r_1, r_2) &= \frac{3r_1^2 + r_2^2}{3r_1} \\ \rho_1(r_1, r_2) &= \frac{r_2^3 - 5r_1^2 r_2}{5r_1^2} \end{aligned}\right\} \qquad r_1 > r_2 .$$

Since

$$J_{3/2}(kr) = \left(\frac{2}{\pi kr}\right)^{1/2}\left(\cos kr - \frac{\sin kr}{kr}\right),$$

the integrations over r_1, r_2 involved in I_A and I_B can be performed without difficulty to give

$$\begin{aligned} I = (2048\pi^3)^{1/2} N\sigma^3 k^{-2} [&6\gamma\rho_3^4(4k\rho_3 \cos 5\xi_1 - \sin 4\xi_1) + 2(\beta - 6\gamma\sigma)\rho_3^3 \\ &(3k\rho_3 \cos 4\xi_1 - \sin 3\xi_1) + (1 + 12\gamma\sigma^2)\rho_3^2(2k\rho_1 \cos 3\xi_1 - \sin 2\xi_1) \\ &- 2\beta\{\sigma\rho_4^2(2k\rho_4 \cos 3\xi_2 - \sin 2\xi_2) + 6\sigma^2\rho_4(k\rho_4 \cos 2\xi_2 - \sin \xi_2) \\ &+ 12\,\sigma^3\rho_4 k \cos \xi_2 - 12a\sigma^4(\xi_1 - \xi_2)\}] , \end{aligned}$$

where

$$\rho_3 = (k^2 + a^2)^{-(1/2)}, \qquad \rho_4 = \{k^2 + (1 + 2a)^2\}^{-(1/2)},$$

$$\xi_1 = \arctan\frac{k}{a}, \qquad \xi_2 = \arctan\frac{k}{1 + 2a},$$

$$\sigma = (1 + a)^{-1} .$$

In Table 1 the numerical values of the continuous absorption coefficient for H^-, calculated for a series of wave numbers of the incident quanta, are given. In the range of importance in astrophysics it will be seen that Q is in the neighborhood of its maximum. Before proceeding to discuss the errors arising from the use of the simple

form in equation (3) for the continuous wave function, it is of interest to compare our results with those obtained previously by Jen[10] and by Massey and Smith.[11] Jen obtains results somewhat similar to ours for the higher frequencies, but his absorption coefficient tends to a finite value at the low-frequency limit, whereas ours tends to zero. This difference arises from our use of a plane wave to represent the behavior of the ejected electron, whereas Jen uses a wave modified by a Coulomb field of potential $-0.4e^2/r$. For small en-

TABLE 1

CALCULATED ABSORPTION COEFFICIENTS FOR H^-

Wave Number of Radiation in 10^3 Cm^{-1}	Absorption Coefficient in 10^{-17} Cm^2	Wave Number of Radiation in 10^3 Cm^{-1}	Absorption Coefficient in 10^{-17} Cm^2
5.674 (head)	0.00	.....	.....
10	0.86	45	1.66
15	1.87	50	1.41
20	2.49	60	1.02
25	2.62	70	0.76
30	2.49	80	0.56
35	2.25	90	0.42
40	1.94	100	0.33

ergies of the ejected electron the major contribution to the integral in equation (1) comes from large distances. At such distances his assumption of a Coulomb interaction between the neutral hydrogen atom and the second electron becomes grossly in error, and he obtains too large a value of Q. On the other hand, our results differ from those of Massey and Smith in the other direction. They find $Q \to 0$ at the low-frequency limit as do we, but their maximum occurs at much higher frequencies. This is due to their use of a rough approximate form for the ground-state wave function, considerably more compact than that given by equation (2).

EVALUATION OF CORRECTING TERMS

As the accurate determination of the continuous wave function $\Psi_E(\mathbf{r}_1, \mathbf{r}_2)$ is a problem essentially the same as that which arises in

[10] *Loc. cit.* [11] *Loc. cit.*

discussing the collisions of electrons with hydrogen atoms, so it is convenient to adopt the method of collision theory.[12] We write

$$\Psi_E(\mathbf{r}_1, \mathbf{r}_2) = \{\Omega(\mathbf{r}_1, \mathbf{r}_2) + \Omega(\mathbf{r}_2, \mathbf{r}_1)\}\, 2^{-(1/2)}, \tag{4}$$

where $\Omega(\mathbf{r}_1, \mathbf{r}_2)$ has the asymptotic form for large r_1

$$e^{ikz_1}\chi_{10}(\mathbf{r}_2) .$$

The expression $\Omega(\mathbf{r}_1, \mathbf{r}_2)$ must satisfy the wave equation

$$\left[\frac{h^2}{8\pi^2 m}(\nabla_1^2 + \nabla_2^2) + E + \frac{e^2}{r_1} + \frac{e^2}{r_2} - \frac{e^2}{r_{12}}\right]\Omega(\mathbf{r}_1, \mathbf{r}_2) = 0 . \tag{5}$$

where E is now the total energy of the system.
As it must be a proper function of $\mathbf{r}_1$, $\mathbf{r}_2$, it may be expanded in a series

$$\Omega(\mathbf{r}_1, \mathbf{r}_2) = (\Sigma_{nl} + \int)\chi_{nl}(\mathbf{r}_2)F_{nl}(\mathbf{r}_1), \tag{6}$$

where the $\chi_{nl}(\mathbf{r}_2)$ are the sequence of wave functions for the different states of the hydrogen atom and so satisfy the equation

$$\left(\frac{h^2}{8\pi^2 m}\nabla_2^2 + E_n + \frac{e^2}{r_2}\right)\chi_{nl} = 0 . \tag{7}$$

The integral sign arises from the contribution of the continuous wave functions of the hydrogen atom. The suffixes *nl* represent the total and azimuthal quantum numbers, respectively, of the discrete hydrogen states, so

$$\chi_{nl}(\mathbf{r}_2) = u_{nl}(r_2)Y_l(\theta_2, \phi_2) ,$$

where Y_l is a spherical harmonic of order l. Expanding

$$F_{nl}(\mathbf{r}_1) = \Sigma(2s + 1)F^s_{nl}(r_1)Y_s(\theta_1, \phi_1),$$

[12] Mott and Massey, *The Theory of Atomic Collisions*, chap. x, Clarendon Press, 1933.

and substituting in equations (6) and (1), we find, since

$$\int\int Y_s Y_{s'} \sin\theta d\theta d\phi = 0 , \qquad s \neq s' ,$$

$$= \frac{4\pi}{2s+1} , \qquad s = s' ,$$

$$\int\int r_1 Y_1(\theta_1, \phi_1)\Phi(\mathbf{r}_1, \mathbf{r}_2)\Psi_E(\mathbf{r}_1, \mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2$$

$$= (12S)^{1/2}\pi^2 N \int_0^\infty \int_0^\infty r_1^3 r_2^2 e^{-a(r_1+r_2)}[(1 + \gamma\overline{r_1 - r_2}^2)\Sigma_n\{u_{n0}(r_2)F_{n0}^1(r_1)$$
$$+ u_{n1}(r_1)F_{n1}^0(r_2)\} + \beta\Sigma_n\Sigma_l\{\rho_l(r_1, r_2)u_{nl}(r_2)$$
$$(\mu F_{nl}^{l+1}(r_1) + \nu F_{nl}^{l-1}(r_1)) + u_{nl}(r_1)(\mu\rho_{l+1}(r_1, r_2)F_{nl}^{l+1}(r_2)$$
$$+ \nu\rho_{l-1}(r_1, r_2)F_{nl}^{l-1}(r_2))\}]dr_1 dr_2 + I_c ,$$

where μ, ν are numerical coefficients

and

$$r_{12} = \Sigma(2s+1)\rho_s(r_1, r_2)P_s(\cos\odot_{12}) .$$

I_c represents the contribution from the terms of $\Psi_E(\mathbf{r}_1, \mathbf{r}_2)$ involving the continuous wave functions of the hydrogen atom. It is apparent that only certain terms in the expansions of the functions F_{nl} give nonvanishing contributions to the absorption coefficient, and this simplification is important. To determine the functions F_{nl} we substitute equation (6) in equation (5) and use equation (7) to give

$$(\Sigma_{nl} + \int)\chi_{nl}(\mathbf{r}_2)\left\{\frac{h^2}{8\pi^2 m}\nabla_1^2 + E - E_n\right\}F_{nl}(\mathbf{r}_1) = \left(\frac{e^2}{r_{12}} - \frac{e^2}{r_1}\right)\Omega(\mathbf{r}_1, \mathbf{r}_2) . \quad (8)$$

Multiplying both sides of this equation by $\chi_{nl}^*(\mathbf{r}_2)$ and integrating over $\mathbf{r}_2$ gives, on account of the orthogonal properties of $\chi_{nl}(\mathbf{r}_2)$,

$$\left\{\frac{h^2}{8\pi^2 m}\nabla_1^2 + E - E_n\right\}F_{nl}(\mathbf{r}_1) = e^2\int\left(\frac{1}{r_{12}} - \frac{1}{r_1}\right)\Omega(\mathbf{r}_1, \mathbf{r}_2)\chi_{nl}^*(\mathbf{r}_2)d\tau_2 . \quad (9)$$

In view of the form of equation (6) for Ω, this gives an infinite set of simultaneous integro-differential equations for the functions F_{nl}. Solutions are required which have the asymptotic form for large r_1,

$$F_{10}(\mathbf{r}_1) \sim e^{ikz_1} + r_1^{-1}e^{ikr_1}f_{10}(\theta_1, \phi_1) ,$$

$$F_{nl}(\mathbf{r}_1) \sim r_1^{-1}\, e^{ik_n r_1}f_{nl}(\theta_1, \phi_1) , \qquad (n, l \neq 1, 0) ,$$

where

$$k_n^2 = \frac{8\pi^2 m}{h^2}(E - E_n) .$$

The functions $F_{nl}(\mathbf{r}_1)$ correspond to electron waves scattered after exciting the atom to the state *nl*. In the energy range of importance in our problem the ejected electrons have insufficient energy to produce excitation. This does not mean that all the $F_{nl}(\mathbf{r}_1)$ vanish, but they have the asymptotic form

$$r_1^{-1} e^{-\gamma_n r_1} f_{nl}(\theta_1, \phi_1) ,$$

where $\gamma_n^2 = -k_n^2$.

Terms of the form

$$\chi_c(\mathbf{r}_2) F_c(\mathbf{r}_1) ,$$

where $\chi_c(\mathbf{r}_2)$ is a continuous hydrogen wave function, represent capture of the outgoing electron and ejection of the atomic electron. This corresponds to electron exchange, and it is possible to take such terms into account directly without explicit introduction of the cumbersome continuous hydrogen wave functions. For convenience we shall refer henceforward to the term of Ω involving $F_{10}(\mathbf{r}_1)$ as the bound-free term, those of form $\chi_c(\mathbf{r}_2)F_c(\mathbf{r}_1)$ as the free-bound terms, and the remainder as bound-bound. Then we can say that the bound-free term includes the unperturbed plane wave and the modification of this wave by the atomic field (including polarization), the free-bound terms represent exchange phenomena, and the bound-bound arise from the nonseparability of Ω.

THE BOUND-BOUND TERMS

Our approximation in equation (3) amounts to taking

$$F_{nl}(\mathbf{r}_1) = 0 ; \qquad n, l \neq 1, 0 ; \qquad F_{10}(\mathbf{r}_1) = e^{ikz_1} .$$

To obtain the next approximation we substitute these values on the right-hand side of equation (9) giving

$$(\nabla^2 - \gamma_n^2) F_{nl}(\mathbf{r}_1) = \frac{8\pi^2 m e^2}{h^2} \int \left(\frac{1}{r_{12}} - \frac{1}{r_1} \right) e^{ikz_1} \chi_{10}(\mathbf{r}_2) \chi_{nl}^*(\mathbf{r}_2) d\tau_2 ; \quad (10)$$

so, using Green's function, we should have

$$F_{nl}(\mathbf{r}) = \frac{2\pi me^2}{h^2}\int\int \frac{\exp(-\gamma_n|\mathbf{r}-\mathbf{r}_1|)}{|\mathbf{r}-\mathbf{r}_1|}\, e^{ikz_1}\left(\frac{1}{r_1}-\frac{1}{r_{12}}\right)\chi_{10}(r_2)\chi^*_{nl}(\mathbf{r}_2)d\tau_1 d\tau_2 . \quad (11)$$

Now, it is well known from the theory of collisions that the F_{nl} are largest for $n, l = 2,1$. To obtain an estimate of the error involved by neglect of the bound-bound terms we therefore proceed to evaluate F_{21} from equation (11), using

$$\chi_{21}(r_2) = (32\pi)^{-(1/2)} r_2 e^{-(1/2)r_2} \begin{cases} \sin\theta_2 \cos\phi_2 \\ \cos\theta_2 \\ \sin\theta_2 \sin\phi_2 . \end{cases}$$

As only the zero-order and second-order harmonics in the expansion of F_{21} are required, the calculation is not difficult. Use is made of the expansion

$$(|\mathbf{r}_1-\mathbf{r}_2|)^{-1} e^{-\gamma|\mathbf{r}_1-\mathbf{r}_2|} = (r_1 r_2)^{-(1/2)} \Sigma (2n+1) P_n\left(\frac{\mathbf{r}_1\cdot\mathbf{r}_2}{r_1 r_2}\right)\eta_n ,$$

where

$$\eta_n = K_{n+1/2}(\gamma r_1) I_{n+1/2}(\gamma r_2); \qquad r_1 > r_2 .$$

$K_{n+1/2}, I_{n+1/2}$ are Bessel functions which can be obtained in closed form so the integration may be carried out analytically except for integrals of the form

$$\int^r \begin{matrix}\cos \\ \sin\end{matrix} kr\; e^{\pm ar} r^{-1} dr ,$$

which must be evaluated numerically. The result of the calculation shows that the effect of the bound-bound terms is not greater than 1 per cent of the total.

THE BOUND-FREE AND FREE-BOUND TERMS

Although the method of successive approximation is satisfactory in evaluating F_{nl} where $nl \neq 1, 0$, it is well known in the theory of collisions that it cannot be relied upon to estimate either the second approximation to F_{10} (the effect of distortion of the plane wave by the atomic field) or the effect of exchange. This is primarily due to the fact that, with $n, l \neq 1, 0$ the magnitude of F_{nl} is determined by

the "transition" potentials which are nondiagonal matrix elements of the interaction. The correction to F_{10} involves the usually much larger diagonal matrix element. To avoid this difficulty, however, one may proceed by ignoring all terms in Ψ_E which do not involve waves in either r_1 or r_2, of the same wave length as the ejected electron. That is to say, we write

$$\Psi_E(\mathbf{r}_1, \mathbf{r}_2) = 2^{-(1/2)}\{\chi_{10}(r_2)F_{10}(\mathbf{r}_1) + \chi_{10}(r_1)F_{10}(\mathbf{r}_2)\}$$

and substitute in equation (5) directly. As has been shown by Morse and Allis,[13] this leads to the equation

$$(\nabla^2 + k^2 - v)F_{10}(\mathbf{r}_1) = u\chi_{10}(r_1) \tag{12}$$

where

$$v = \frac{8\pi^2 me^2}{h^2}\int\left(\frac{1}{r_1} - \frac{1}{r_{12}}\right)\chi_{10}(r_2)\chi^*_{10}(r_2)d\mathbf{r}_2$$

and

$$u = \frac{8\pi^2 me^2}{h^2}\left\{2\int\chi_{10}(r_2)F_{10}(\mathbf{r}_2)\frac{d\mathbf{r}_2}{r_{12}} + (k^2 - E_1)\int\chi_{10}(r_2)F_{10}(\mathbf{r}_2)d\mathbf{r}_2\right\}.$$

The term v is the potential of the undisturbed atomic field and u is the additional term due to exchange.

In our case $F^1_{10}(r_1)$ is alone of interest and for it we may write the equation in the form

$$\left(\frac{d^2}{dr_1^2} - \frac{2}{r_1^2} - v + k^2\right)r_1F^1_{10}(r_1) = u_1\chi_{10}(r_1)\,, \tag{13}$$

where

$$u_1 = \frac{64\pi^3 me^2}{3h^2}\left\{r_1^2\int_{r_1}^{\infty}\chi_{10}(r_1)F^1_{10}(r)r^{-2}dr + r_1^{-1}\int_0^{r_1}\chi_{10}(r)F^1_{10}(r)rdr\right\}.$$

Now, if the simple approximation in equation (3) is valid, we must show that the substitution

$$r_1F^1_{10}(r_1) = \cos kr_1 - \frac{\sin kr_1}{kr_1}\,,$$

is really a very good approximation to the solution of equation (13), i.e., we must show that, writing V for $v + u_1\chi_{10}(r_1)/r_1F^1_{10}(r_1)$,

$$\frac{2}{r^2} \gg V\,.$$

[13] *Phys. Rev.*, **44**, 269, 1933.

In Figure 1 the relative magnitudes of the terms involved are illustrated, and it is quite clear that our simple approximation is a very satisfactory one. The effect of exchange is to improve the accuracy of the approximation as it reduces the effective atomic field. As a final test of the validity of equation (3), we solved equation (13) exactly, with exchange neglected, thus obtaining an upper limit to

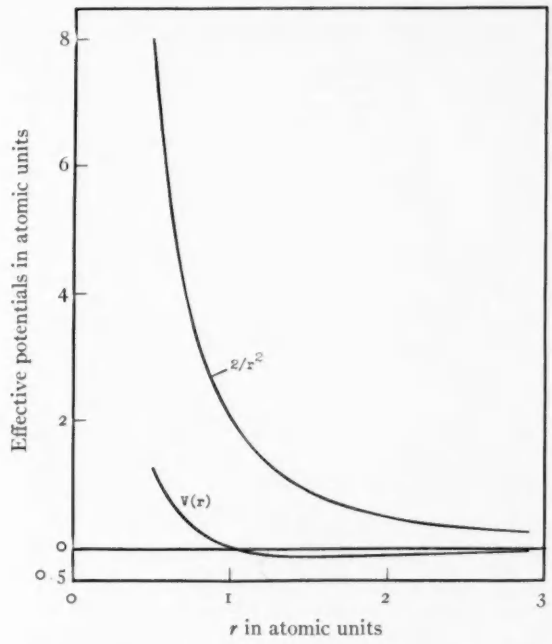


FIG. 1.—Illustrating the relative magnitude of the terms $2/r^2$ and $V(r)$ which determine the form of the wave function of the ejected electron.

the correction necessary. It was found that no error greater than 4 per cent arises from neglect of exchange and distortion of the plane wave by the undisturbed atomic field (v).

The effect of polarization of the atomic field on the function $F^1_{10}(r_1)$ is much smaller still, as it arises only in the higher approximations. Thus, substituting equation (6) for Ω on the right-hand side of equation (9), we find in the equation for $F_{10}(r_1)$ an additional effective potential

$$\Sigma V_{10,\,nl}(r_1)F_{nl}(r_2)\ ,$$

where

$$V_{10,\,nl} = e^2 \int \left(\frac{1}{r_{12}} - \frac{1}{r_1} \right) \chi_{10}(r_2) \chi^*_{nl}(r_2) d\tau_2 \,.$$

The largest contribution to this will again come from n, $l = 2$, 1, and it was estimated to be less than 10 per cent of the exchange and distortion corrections.

It is of interest to notice here that if the zero-order harmonic of F_{10} had been required instead of the first order, the problem of accurately determining it would have been much more difficult. For then we would have $l = 0$, the dominant term $l(l+1)/r^2$ of equation (13) would vanish, and a large perturbation from the atomic field and exchange would result (possibly also polarization effects would be important).

It is therefore established that our approximation in equation (3) for Ψ_E is never likely to introduce an error of more than a few per cent at most in our calculated absorption coefficient. As we have given reasons to expect that the discrete wave function is also to be relied upon, at least in so far as it determines Q in the region of importance in astrophysics, the values given for the absorption coefficient can be regarded as accurate to a much higher degree than in any previous calculation of continuous absorption coefficients by atoms or ions with more than one electron. In no other cases have the effects of exchange, polarization, and nonseparability on the continuous wave function been discussed, and the fact that they may be neglected for H^- by no means insures that they can be neglected for other elements.

We are much indebted to Dr. Strömgren for bringing the problem of H^- to our attention and we have benefited considerably from correspondence with him and with Dr. Wildt, to whom we wish also to express our thanks. Our thanks are due moreover to the Ministry of Education for Northern Ireland for a grant to one of us (D. R. B.).

MATHEMATICS DEPARTMENT
UNIVERSITY COLLEGE
LONDON

METASTABILITY OF HYDROGEN AND HELIUM LEVELS

G. BREIT AND E. TELLER

ABSTRACT

The mean life of the 2s state of hydrogen in the absence of collisions is primarily determined by the simultaneous emission of two photons. Owing to this cause the mean life is $\sim\frac{1}{7}$ sec. It is expected that the mean life of 1s 2s ^{1}S of *He* I is of the same order of magnitude and is much shorter than that of 1s 2s ^{3}S of *He* I. The non-adiabatic collisions with electrons are found to be more important than the static effects at small electron densities, leading to a transition probability $\sim_{8}{}^{1}_{0}{}^{1}_{0}$ sec^{-1} for an electron temperature $\sim$10,000° C and an electron density 30 cm^{-3}. The effect of static electric fields on the 2s state of hydrogen is worked out in more detail than previously by taking into account the hyperfine structure. The mean life in static electric fields which are strong enough to make double emission of little importance and yet weak enough to give a mean life $>10^{-7}$ sec is found to be the sum of two lives: (1) the mean life calculated neglecting hyperfine structure and (2) the mean life calculated neglecting radiation coupling. In both (1) and (2) the mixing of 2p into 2s by the Stark effect is taken into account.

Radiations due to quadrupole and magnetic dipole effects are considered qualitatively for 1s 2s ^{3}S of helium and quantitatively for 2s of hydrogen. For the latter there exists a small radiative magnetic dipole effect on Dirac's theory of the electron which is absent in the ordinary spin model.

INTRODUCTION

Struve, Wurm, and Henyey[1] consider the mean life of the metastable level of atomic hydrogen (2s) using Bethe's calculation[2] of the influence of the electric field. At the very low pressures of interstellar space the mean life exceeds the values obtainable in the laboratory by such a large factor that it becomes necessary to consider some additional effects. There are in the first place dynamic effects of electrons. The collision cross-section for the transfer from the metastable level to the $2p_{1/2}$ level gives a transition probability of the order of $5 \times 10^{-5}n$ sec^{-1} at an absolute electron temperature of 10,000°. Here n is the number of electrons per cm^3. The possible importance of this process was mentioned to us first by Professor E. Wigner. The static effect of the electric field is smaller at the supposed values of T and n. It is considered somewhat in detail on account of the in-

[1] *Proc. Nat. Acad.*, **25**, 67, 1939; see also O. Struve and K. Wurm, *Ap. J.*, **88**, 84, 1938, for discussion of electron collisions in *He*, and O. Struve, *Proc. Nat. Acad.*, **25**, 36, 1939, for electron temperature.

[2] *Handbuch d. Phys.* (Geiger-Scheel), **24**, Part 1, 452, 1933.

teresting physical problem presented by the overlap of the hyperfine-structure pattern by the radiation width of the $2p_{1/2}$ level.

Second, the simultaneous emission of two photons[3] makes it impossible for the 2s level to have a life longer than $\sim \frac{1}{10}$ sec even in the absence of effects of collisions and of electric fields. Double emission is presumably also important for the 1s 2s ^1S state of *He* I in which it may be expected to be of the same order of magnitude as for hydrogen. For 1s 2s ^3S of *He* I, however, selection rules make double emission less probable. We understand from Professor Struve that there is some astrophysical evidence for a shorter mean life of 1s 2s ^1S as compared with that of 1s 2s ^3S.

I. EFFECTS OF ELECTRONS

A. THE EFFECT OF THE STATIC ELECTRIC FIELD

It will be found later that the dynamic effects are more important than the static ones. In the interest of system in presentation, static effects will be considered first and the effect of hyperfine structure will be worked out. From a hasty examination of the problem it would appear that the effect of electric fields may be quite different with and without hyperfine structure (hfs) because the accidental degeneracy of the $2s_{1/2}$, $2p_{1/2}$ states is removed by the nuclear moment. This is not so, however, because the width of the $2p_{1/2}$ level due to radiation broadening covers appreciably the hfs pattern.

Even though the average electric field assumed to exist statistically in interstellar space causes a practically negligible distortion of the hyperfine-structure pattern, its effect on the average mean life turns out to be relatively important. For small fields the mean life for a given value of the fine-structure magnetic quantum number will be seen to be the sum of two parts. The first of these is the mean life due to the admixture of the $2p_{1/2}$ state caused by the electric field and obtainable by neglecting the effect of coupling between stationary levels due to their natural width. The second contribution is due to the radiation breadth of the $2p_{1/2}$ level which diminishes the admixture of $2p_{1/2}$ to 2s and increases the mean life. This increase appears in the result as an additive term, provided the electric field is weak.

[3] Maria Goeppert, *Naturwiss.*, **17**, 932, 1929; Maria Goeppert Mayer, *Ann. d. Phys.*, **9**, 273, 1931.

Assuming the proton magnetic moment to be 2.8 nuclear magnetons, the hyperfine structure of the $2s_{1/2}$, $2p_{1/2}$ levels is shown in Figure 1. The hyperfine-structure quantum number is denoted by f

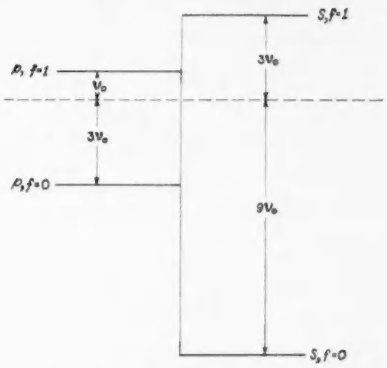


FIG. 1.—Hyperfine-structure pattern of the 2s, $2p_{1/2}$ group of levels. The dashed line represents position of the group in the absence of hyperfine structure. Frequency unit $\nu_0/c = 0.00049\ \text{cm}^{-1}$.

in the figure and below. It is equal to the angular momentum of electron+proton in units $\hbar$. The $s_{1/2}$, $p_{1/2}$ levels are split by the interaction with the nucleus into components having a separation of 0.006 cm^{-1} for the first and 0.002 cm^{-1} for the second.

The average electric field used is

$$\left.\begin{aligned} \mathfrak{E} &= \frac{e}{l^2}, \\ l &= \frac{1}{2}\left(\frac{3}{4\pi n}\right)^{1/3} = \frac{n^{-1/3}}{3.22}, \end{aligned}\right\} \tag{1}$$

where n = number of electrons per cm^3. The Stark effect splitting caused by this field in the absence of hyperfine structure is

$$2\Delta w = 4(3)^{1/2}\left(\frac{a_H}{l}\right)^2 \frac{e^2}{2a_H}$$

$$\frac{2\Delta w}{hc} = 7.5 \times 10^5 \left(\frac{a_H}{l}\right)^2 \text{cm}^{-1}. \tag{2}$$

Here $e^2/2a_H$ expressed in wave numbers is 1.09×10^5 cm^{-1} and a_H is the Bohr radius. For $n = 30$ cm^{-3} one has $l = 0.10$ cm and

$2\Delta w/hc = 2 \times 10^{-9}$ cm^{-1}. This is very much smaller than the separations between hyperfine-structure levels.

The mean life of the 2p level of hydrogen is 1.6×10^{-9} sec corresponding to a half-value breadth of 10^8 sec^{-1} in frequency units or 0.003 cm^{-1} in wave numbers. This breadth is sufficient to cover the hyperfine components of the 2s level by the natural width of the $2p_{1/2}$ level.

The mean life under the foregoing conditions can be treated by the method of the damping matrix.[4] This method considers the emission and absorption from groups of atomic levels separated by energy intervals of the order of magnitude of the radiation width. The photons emitted when the atom leaves the level b' may be reabsorbed so that a neighboring level b'' becomes populated. An exponential decay of b' occurs only in conjunction with an exponential decay of b'' having the same time constant. Definite mean lives do not belong, therefore, to b' or b'' but rather to linear combinations of the atomic states. The calculations concern themselves with finding the linear combinations and the mean lives.

The probability amplitudes of the atomic states b' are denoted by $c(b')$. For solutions corresponding to definite mean lives

$$c(b') = \sum(\bar{b})A(b', \bar{b})a(\bar{b}) . \tag{3}$$

A definite mean life corresponds to any given $\bar{b}$. The state corresponding to it is obtained by setting $a(\bar{b}) = 1$ and $a(\bar{b}') = 0$ for $\bar{b}' \neq \bar{b}$. The probability amplitudes for such a state are $c(b') = A(b', \bar{b})$. Their values can be determined by solving

$$A(b', \bar{b})[\omega(b') - \omega(\bar{b}) + i\gamma(\bar{b})] = i\sum(b'')\gamma(b', b'')A(b'', \bar{b}) , \tag{3.1}$$

where $\omega(b') = 2\pi\nu(b')$ and ν stands for the transition frequency to the ground level corresponding to the level b'. The $\gamma(b', b'')$ are matrix elements of the damping matrix and can be computed by

$$\left.\begin{aligned}\gamma(b', b'') = \gamma_0\sum(a')[&x(b', a')x(a', b'') + y(b', a')y(a', b'') \\ &+ z(b', a')z(a', b'')]\end{aligned}\right\} \tag{3.2}$$

and

$$\gamma_0 = \frac{32\pi^4e^2\nu^3}{3hc^3} , \tag{3.3}$$

[4] G. Breit, *Rev. Mod. Physics*, **5**, 91, 1933 (see pp. 118–125).

the summation being performed over the lower levels a' located outside the group of interacting levels b'. Direct transitions between b' due to single photon emissions and absorptions are here neglected. The complex numbers $\omega(\bar{b}) - i\gamma(\bar{b})$ are determined by the requirement of compatibility of the set of linear equations for $A(b', \bar{b})$ similarly to the conditions for real characteristic values. The ratios of the $A(b', \bar{b})$ for fixed $\bar{b}$ and variable b' determine that linear combination of atomic states which has a time dependence

$$exp\,[-i\omega(\bar{b})t - \gamma(\bar{b})t]$$

for which the mean life is

$$\tau(\bar{b}) = \frac{1}{2\gamma(\bar{b})}\,. \tag{3.4}$$

The general theory will now be applied to the special case of the 2s, 2p group of hydrogen levels split by hyperfine structure and in a weak electric field. The magnetic quantum number of the hyperfine-structure level (in the absence of an electric field) will be called μ, and the stationary states corresponding to given f and μ will be called $\psi_{sf\mu}$ and $\psi_{pf\mu}$ for s and p states, respectively. In an electric field the stationary states of the atomic system are

$$\left.\begin{aligned} \varphi_{s\mu} &= c(\mathrm{s,\,s})\psi_{sf\mu} + c(\mathrm{s,\,p})\psi_{pf'\mu} \\ \varphi_{p\mu} &= c(\mathrm{p,\,s})\psi_{sf\mu} + c(\mathrm{p,\,p})\psi_{pf'\mu}\,, \end{aligned}\right\} \tag{4}$$

where $c(\mathrm{s,\,s})$, $c(\mathrm{s,\,p})$, etc., are coefficients of a unitary transformation. Since the electric field is weak, $c(\mathrm{s,\,s})$ and $c(\mathrm{p,\,p})$ are nearly unity while $c(\mathrm{s,\,p})$ and $c(\mathrm{p,\,s})$ are nearly zero. The linear combinations in equations (4) contain two wave functions only, because the electric field directed along the z axis is a matrix diagonal in μ. There are four sets of equations (4) corresponding to $(f,\ \mu) = (1,\ 1)$, $(1,\ 0)$, $(1,\ -1)$, $(0,\ 0)$. The matrix elements $\gamma(b', b'')$ must be referred to the states φ since these are the stationary states of the atom. The selection rules of this matrix are simpler when referred to the ψ. In this representation the nonvanishing elements of γ belong to a row and column with the same $\psi_{pf\mu}$, and all these di-

agonal elements are equal to each other. The stationary states represented by equation (4) are

$$\left.\begin{aligned} S_{1\mathfrak{E}} &= A_1 S_1 + B_1 P_1 , & S_{-1\mathfrak{E}} &= A_{-1} S_{-1} + B_{-1} P_{-1} ; \\ P_{1\mathfrak{E}} &= -B_1 S_1 + A_1 P_1 , & P_{-1\mathfrak{E}} &= -B_{-1} S_{-1} + A_{-1} P_{-1} ; \\ S'_{0\mathfrak{E}} &= A' S_0 + B' P^0_0 , & S''_{0\mathfrak{E}} &= A'' S^0_0 + B'' P^0_0 ; \\ P'_{0\mathfrak{E}} &= -B' S_0 + A' P^0_0 , & P''_{0\mathfrak{E}} &= -B'' S^0_0 + A'' P^0_0 . \end{aligned}\right\} \quad (4.1)$$

Here

$$S_1 = \psi_{s11} , \qquad S_0 = \psi_{s10} , \qquad S_{-1} = \psi_{s1-1} , \qquad S^0_0 = \psi_{s00}$$

and similarly for P_1, etc. The elements of γ are zero except in the squares determined by $(S_{1\mathfrak{E}}, P_{1\mathfrak{E}})$, $(S_{-1\mathfrak{E}}, P_{-1\mathfrak{E}})$, $(S'_{0\mathfrak{E}}, P'_{0\mathfrak{E}})$, and $(S_{0\mathfrak{E}}, P''_{0\mathfrak{E}})$. The wave functions ψ, φ will be chosen so that all A, B in the foregoing equations are real. The nonvanishing elements of γ are

$$\left.\begin{aligned} \gamma(S_{\mathfrak{E}}, S_{\mathfrak{E}}) = 2\gamma B^2 ; \qquad \gamma(S_{\mathfrak{E}}, P_{\mathfrak{E}}) = \gamma(P_{\mathfrak{E}}, S_{\mathfrak{E}}) = 2\gamma AB ; \\ \gamma(P_{\mathfrak{E}}, P_{\mathfrak{E}}) = 2\gamma A^2 , \end{aligned}\right\} \quad (4.2)$$

where indices and primes are omitted. The constant γ is related to the mean life of the 2p level in the absence of the electric field by

$$\tau_{2p} = \frac{1}{4\gamma} . \quad (4.3)$$

In a strong electric field $A^2 = B^2 = \frac{1}{2}$, and the mean life of all levels is $\frac{1}{2}\gamma$. For intermediate fields equation (3.1) gives

$$\left.\begin{aligned} \omega(\bar{b}) - i\gamma(\bar{b}) = -i\gamma + \tfrac{1}{2}[\omega(S_{\mathfrak{E}}) + \omega(P_{\mathfrak{E}})] \\ \pm \{[\tfrac{1}{2}\omega(S_{\mathfrak{E}}) - \tfrac{1}{2}\omega(P_{\mathfrak{E}}) + i\gamma]^2 - 2i\gamma B^2[\omega(S_{\mathfrak{E}}) - \omega(P_{\mathfrak{E}})]\}^{1/2} . \end{aligned}\right\} \quad (5)$$

The $\omega(S_{\mathfrak{E}})$, $\omega(P_{\mathfrak{E}})$, B^2 can be calculated by standard methods, and hence $\tau(\bar{b})$ is determined by equation (3.4). For weak fields B varies linearly with $\mathfrak{E}$. Including effects of order $\mathfrak{E}^2$ only, it suffices to expand the foregoing formula including only the term in B^2. The minus sign of the radical gives $\gamma(\bar{b}) = 2\gamma +$ term in B^2. For vanishing $\mathfrak{E}$ this solution has the mean life τ_{2p}, and the correction due to the term in $\mathfrak{E}^2$ is negligible for the small $\mathfrak{E}$ under consideration. The $A(b', \bar{b})$ corresponding to this root is small for the coefficient of φ_s and nearly

1 for φ_p. This solution corresponds to the atom's being primarily in the p state during the emission. The plus sign of the radical gives $\gamma(\bar{b}) = 0$ for $\mathfrak{E} = 0$. This root gives $A(b', \bar{b})$ equal to unity for φ_s and zero for φ_p. If $\mathfrak{E} \neq 0$, the φ_p and hence also the ψ_p are present only to a small extent. Calculating $\gamma(\bar{b})$ by expanding equation (5), one obtains for small $\mathfrak{E}$

$$\overline{\gamma}(s) = \gamma(\bar{b}) = \frac{2\gamma B^2[\omega(S_{\mathfrak{E}}) - \omega(P_{\mathfrak{E}})]^2}{[\omega(S_{\mathfrak{E}}) - \omega(P_{\mathfrak{E}})]^2 + 4\gamma^2}, \tag{5.1}$$

so that by equation (3.4)

$$\tau_{2s} = \tau_{2p}B^{-2}\{1 + \tfrac{1}{4}\tau_{2p}^{-2}[\omega(S_{\mathfrak{E}}) - \omega(P_{\mathfrak{E}})]^{-2}\}. \tag{5.2}$$

If $\tau_{2p}^2[\omega(S_{\mathfrak{E}}) - \omega(P_{\mathfrak{E}})]^2 \gg 1$, the second term in the braces is negligible, and then $\tau_{2s} = \tau_{2p}B^{-2}$. This is just the result which would be obtained neglecting the overlap of the s level by the natural width of the p level and attributing the mean life to the presence of $B\varphi_p$ in the wave function. The fractional error in τ_{2s} due to the neglect of the second term in braces becomes negligible if $[\omega(S_{\mathfrak{E}}) - \omega(P_{\mathfrak{E}})]^2$ is increased sufficiently. On the other hand, the absolute error due to this term is independent of $[\omega(S_{\mathfrak{E}}) - \omega(P_{\mathfrak{E}})]^2$, because $B^2[\omega(S_{\mathfrak{E}}) - \omega(P_{\mathfrak{E}})]^2$ does not depend on $[\omega(S_{\mathfrak{E}}) - \omega(P_{\mathfrak{E}})]^2$ for small $\mathfrak{E}$. For small $[\omega(S_{\mathfrak{E}}) - \omega(P_{\mathfrak{E}})]^2$ this term gives the main effect, since then $\tau_{2p}B^{-2}$ is negligible. The mean life for small $\mathfrak{E}$ may be thus thought of as

$$\tau = \tau \text{ (electric field mixing)} + \tau \text{ (radiation coupling)}, \tag{5.3}$$

where τ (electric field mixing) $= \tau_{2p}B^{-2}$ and may be computed by neglecting radiation coupling but taking into account the hyperfine structure; τ (radiation coupling) is, on the other hand, independent of hyperfine structure and may be computed assuming the hfs splitting to be zero.

Calculation shows that the four B's of equation (4.1) are given by

$$B^2 = 3\mathfrak{E}^2e^2a_H^2\hbar^{-2}[\omega(S_{\mathfrak{E}}) - \omega(P_{\mathfrak{E}})]^{-2}, \tag{5.4}$$

where a_H is the Bohr radius. In this formula the frequency difference belonging to the pair of interacting levels with which B occurs in equation (4.1) is supposed to be used. The difference between $\omega(S_{\mathfrak{E}}) - \omega(P_{\mathfrak{E}})$ and $\omega(S) - \omega(P)$ is negligible on account of the weakness of the electric field. The values of $\omega(S) - \omega(P)$ are $2\pi \cdot 2\nu_0$

for $(s, f = 1, \mu = +1; p, f = 1, \mu = +1)$ and for $(s, f = 1, \mu = -1; p, f = 1, \mu = -1)$; it is $2\pi \cdot 6\nu_0$ for $(s, f = 1, \mu = 0; p, f = 0, \mu = 0)$ and $-2\pi \cdot 10\nu_0$ for $(s, f = 0, \mu = 0; p, f = 1, \mu = 0)$. The first two levels give short mean lives, and the last two longer ones. Substituting equation (5.4) into equation (5.2), we find

$$\tau_{2s} = \frac{\hbar^2[\omega(S) - \omega(P)]^2}{12\gamma\mathfrak{E}^2e^2a_H^2} + \frac{\hbar^2\gamma}{3\mathfrak{E}^2e^2a_H^2}. \tag{5.5}$$

Here

$$\gamma = 1.57 \times 10^8 \text{ sec}^{-1},$$

and

$$\frac{3\mathfrak{E}^2e^2a_H^2}{\hbar^2} = 12\left(\frac{a_H}{l}\right)^4\left(\frac{e^2}{2a_H\hbar}\right)^2 = 430n^{4/3} \text{ sec}^{-2},$$

where equation (1) is used, and

$$\frac{e^2}{2a_H h} = 3 \times 10^{10} \times 1.10 \times 10^5 \text{ sec}^{-1}.$$

Substituting these numbers, we find the two short mean lives to be

$$\tau'_{2s} = (1.2 + 3.6) \times 10^5 n^{-4/3} \text{ sec} = 4.8 \times 10^5 n^{-4/3} \text{ sec} \tag{5.6}$$

and the two long lives to be

$$\left.\begin{aligned} \tau''_{2s} &= (11 + 3.6) \times 10^5 n^{-4/3} \text{ sec} = 15 \times 10^5 n^{-4/3} \text{ sec} \\ \text{and} \quad \tau'''_{2s} &= (31 + 3.6) \times 10^5 n^{-4/3} \text{ sec} = 35 \times 10^5 n^{-4/3} \text{ sec}. \end{aligned}\right\} \tag{5.7}$$

The first number in parentheses is in each case the contribution due to $\tau_{2p}B^{-2}$. The mean life is seen to be longer than without taking into account the hyperfine structure on the average in the ratio $15:3.6 = 4$.

B. DYNAMIC EFFECTS

The foregoing discussion is incomplete inasmuch as it replaces the actual field by an average value and neglects the time dependence of the field. If the electrons were at rest and the hydrogen atoms were drifting very slowly through the electron atmosphere, the static formulas (5) and (5.5) could be applied directly. Atoms located in regions of small $\mathfrak{E}$ would then have a longer life than those in regions

of large $\mathfrak{E}$. If the number of 2s atoms produced per unit volume per second be p, the equilibrium equation in such a static model is

$$p = N(2s)[\tau_{2s}(\mathfrak{E})]^{-1}$$

where $N(2s)$ is the local density of 2s atoms. The average density of 2s atoms is then

$$\overline{N}(2s) = \frac{\int p\tau_{2s}(\mathfrak{E})dxdydz}{\int dxdydz} = p\overline{\tau}_{2s}(\mathfrak{E}) ,$$

the integrals being extended over a region containing many electrons and atoms. In such a static picture the important quantity is the space average of τ_{2s}. The statistics of the variation of $\mathfrak{E}$ in space have been worked out by Holtsmark and by Verweij.[5] According to Holtsmark the chance of $\mathfrak{E}$ being in $\mathfrak{E}$, $\mathfrak{E} + d\mathfrak{E}$ behaves as $\mathfrak{E}^{-5/2}d\mathfrak{E}$ for large $\mathfrak{E}$ and as $\mathfrak{E}^2 d\mathfrak{E}$ for small $\mathfrak{E}$. Equation (5.5) gives therefore convergent results for $\tau(2s)$, and the difference between the more accurate equation (5) and the approximation (5.5) cannot be very important. The most probable electric intensity is, according to Holtsmark, $\sim 1.5(4.2)^{2/3}n^{2/3}\epsilon \sim 4n^{2/3}\epsilon$ and corresponds crudely to Bethe's consideration which gives equation (1) as used above.

At a temperature of 10,000° the velocity of electrons is $\sim 10^8$ cm/sec, and for $n = 30$ cm^{-3} the distance $l \sim 0.1$ cm. The time during which this distance is covered by the electron is $\sim 10^{-9}$ sec. Attributing roughly the most probable field to the influence of one electron at the distance l, the order of magnitude of the time during which this field lasts is, therefore, $\sim 10^{-9}$ sec which is of the order of magnitude of τ_{2p} and is somewhat smaller than the time $|\nu_{sp}|^{-1}$ corresponding to the transition frequencies between the hyperfine-structure components of 2s and 2p. The latter times range from 3×10^{-8} sec for the interval $2\nu_0$ to 6×10^{-9} sec for the interval $10\nu_0$. The effect of admixing the 2p level by the quadratic Stark effect cannot be fully developed, therefore, at this T and n, although it should become more important at smaller T and n. At larger densities and temperatures the most probable field lasts only a fraction of τ_{2p} without changing its value appreciably. The

[5] J. Holtsmark, *Ann. d. Phys.*, **58**, 577, 1919, and Verweij, *Pub. Astron. Inst. Amsterdam*, No. 5, 1936.

static effects should be noticeably decreased in this case. In fact, for an electric field alternating sinusoidally with the time, the transition probability, neglecting hyperfine structure, is proportional to

$$\overline{\mathfrak{E}^2}\,\frac{2\tau_{2p}}{1+(4\pi\nu\tau_{2p})^2},$$

where ν is the frequency of the field. For high electron densities and temperatures $(4\pi\nu\tau_{2p})^2 \gg 1$ and the equations (5) cannot be applied. There is perhaps an offsetting influence in the statistical character of the time dependence of $\mathfrak{E}$ which occasionally permits the field to last at $\sim e/l^2$ for relatively long times. These effects are, however, difficult to estimate. It is doubtful that they can affect appreciably the decrease in the transition probability due to the alternating character of $\mathfrak{E}$ which has just been discussed.

In addition to effects due to electrons at a distance $\sim l$ which vary with $n^{4/3}$ there are also definitely non-adiabatic effects due to collisions with electrons. These effects give a transition probability proportional to n and will be discussed next. Their possible importance has been mentioned to us in this connection by E. Wigner. For *He* the effect of electron collisions has been considered by Struve and Wurm.[1] In hydrogen this effect is relatively more important because of the proximity of 2s to 2p. Even a comparatively distant collision transfers the atom from the 2s into the 2p state with an appreciable probability. From this state it radiates with the transition probability $(\tau_{2p})^{-1}$ either after or during the collision.

The process is very similar to that considered in the theory of the loss of energy of alpha particles in going through matter, which has been treated by Bohr, Mott, Williams, Bloch, Bethe, and others.[6] The present problem is characterized by the extreme smallness of the excitation energy and by the fact that the incident particle has a velocity comparable to that of the atomic electron. The impact parameter method can be expected according to Mott and Williams to lead to much the same result as the Born method and will be

[6] N. Bohr, *Phil. Mag.*, **25**, 10, 1913; N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions*, chaps. xi, xiii, Oxford University Press, 1933; N. F. Mott, *Proc. Camb. Phil. Soc.*, **27**, 553, 1931; F. Bloch, *Zs. f. Phys.*, **81**, 363, 1933, and *Ann. d. Phys.*, **5**, 285, 1933; H. Bethe, *Ann. d. Phys.*, **5**, 325, 1930; E. J. Williams, *Proc. R. Soc.*, **139**, 163, 1933.

used here since the calculation can be arranged to follow Bohr's classical treatment rather closely. The influence of the electronic and nuclear spin will be neglected at first but will be taken into account later on. The probability that after a collision the atom has been transferred into the p state is then

$$\frac{e^2}{\hbar^2}\,|x_{sp}|^2\left[\left|\int_{-\infty}^{+\infty}\mathfrak{E}_x e^{i\omega t}dt\right|^2+\left|\int_{-\infty}^{+\infty}\mathfrak{E}_y e^{i\omega t}dt\right|^2\right]=C\,,$$

where $\mathfrak{E}_x$, $\mathfrak{E}_y$ are the components of $\mathfrak{E}$ produced at the atom by the colliding particle, taken perpendicular and parallel to the velocity. The frequency $\omega/2\pi$ is the transition frequency between the s and p levels while x_{sp} is the matrix element of the x co-ordinate of the atomic electron between the s level and the p sublevel directed along x. The electron velocity will be called v, and the distance from the atom to the electron path will be called ρ. One finds then

$$\int_{-\infty}^{+\infty}\mathfrak{E}_x e^{i\omega t}dt=\frac{e}{\rho v}\int_{-\infty}^{+\infty}\frac{\cos\dfrac{\omega\rho x}{v}}{(1+x^2)^{3/2}}\,dx=2f\left(\frac{\omega\rho}{v}\right)$$

and

$$\int_{-\infty}^{+\infty}\mathfrak{E}_y e^{i\omega t}dt=\frac{e}{\rho v}\int_{-\infty}^{+\infty}\frac{\sin\dfrac{\omega\rho x}{v}}{(1+x^2)^{3/2}}\,dx=-2g\left(\frac{\omega\rho}{v}\right),$$

where the functions f, g are those introduced by Bohr. The collision cross-section for excitation is

$$\sigma=\int_{\rho_0}^{\infty}C\cdot 2\pi\rho d\rho=\frac{8\pi e^4|x_{sp}|^2}{\hbar^2 v^2}\int_{\omega\rho_0/v}^{\infty}t^{-1}\mathrm{P}(t)dt\,,$$

where

$$\mathrm{P}(z)=f^2(z)+g^2(z)\,,$$

again in Bohr's notation. The lower limit of integration will be taken as

$$\rho_0=\frac{\hbar}{mv}\,.$$

This choice is somewhat arbitrary, but it is natural since for it $C \sim |x_{sp}/a_H|^2$ so that the approximations used in deriving C cease to apply for smaller ρ. The results are not very sensitive to the choice for ρ_0, since ρ_0 enters logarithmically. The effect of collisions with $\rho < \rho_0$ will be neglected. Even though these collisions should be considered for the stopping-power of an atom on account of the large energy transfer, they are not of special importance in the present problem, since only the very small energy transfers are of interest. Proceeding as in Bohr's paper and using his evaluation of his constant $k = 1.12$, one obtains

$$\sigma = \frac{8\pi e^4 |x_{sp}|^2}{\hbar^2 v^2} ln \left(\frac{1.1\, mv^2}{\hbar\omega} \right). \qquad (S)$$

The quantity in the logarithm cannot be considered as certain on account of the arbitrariness in the choice of ρ_0. The factor multiplying the logarithm is just twice that obtained for the excitation of an atom in the normal state by a high-velocity electron. This formula is

$$\sigma = \frac{4\pi e^4 |x_{sp}|^2}{\hbar^2 v^2} ln \left(\frac{2mv^2}{\Delta E} \right), \qquad (F)$$

where ΔE = excitation energy. For a fast electron the wave length is $\ll a_H$, and the integration should be limited at $\sim a_H$ since otherwise the electron penetrates into the atom. In the foregoing calculation modified for fast electrons the quantity under the logarithm is, therefore, essentially $v/(\omega a_H) = vme^2/(\omega\hbar^2)$; for $\omega\hbar \sim e^2/a_H$ this is

$$\frac{mva_H}{\hbar} = \frac{v\hbar}{e^2} = \left(\frac{mv^2 a_H}{e^2} \right)^{1/2} \sim \left(\frac{mv^2}{\Delta E} \right)^{1/2}$$

so that

$$ln \left(\frac{v}{\omega a_H} \right) \sim \tfrac{1}{2} ln \left(\frac{mv^2}{\Delta E} \right).$$

In this way the use of a_H for the lower cutoff together with the relation $\omega\hbar \sim e^2/a_H$ changes the factor 8 in formula (S) to the factor 4 in formula (F). In the present problem $\omega\hbar \ll e^2/a_H$ and $\hbar/mv \sim 10^{-6} T^{-1/2}$ cm. For $T < 10{,}000^\circ$, $\hbar/mv > 10^{-8}$ cm, so that

the condition of the incident electron staying out of the atom is satisfied partially. At temperatures above 10,000° formula (*S*) becomes poor because the velocity of the incident electron is of the order of magnitude of the velocity of the atomic electron, and ρ_0 penetrates into the atom.

A systematic description of the Born method is given by Mott and Massey in *The Theory of Atomic Collisions* (Oxford University Press). On pages 176 and 177 the relation to the present results can be seen. The factor $\frac{1}{2}$ is introduced essentially in their inequality (38) which limits the momentum change to $[2m|E_0|/\hbar^2]^{1/2} = K_0$. If, however, the calculation leading to their equation (40) is made for a case in which one may set $e^{iKx} = 1 + iKx$ for the computation of $\int e^{iKx}\psi_0\psi_n^* d\tau$, then one obtains from equation 13, page 161, essentially twice the value of Q_{0n} as given by equation (40). From this point of view the essential condition for preferring equation (*S*) to equation (*F*) is that $mv < mc/137$ or $v < 2 \times 10^8$ cm/sec. This is also the result of the consideration with the impact parameter method.

Neither equation (*S*) nor equation (*F*) is exactly applicable. For a more exact discussion one would have to evaluate $\int e^{iKx}\psi_s\psi_p d\tau$. For present applications formula (*S*) is probably good enough since the primary question is one of order of magnitude. In view of the fact that the impact parameter method and the Born method indicate that the true result is between formulas (*S*) and (*F*), the former will be used, and the results will be considered in the nature of an overestimate for the transition probability.

On account of the splitting of the 2p level into $2p_{1/2}$, $2p_{3/2}$ and the further splitting of these levels into hyperfine-structure components, the calculation leading to formula (*S*) must be modified. The procedure is straightforward and leads to

$$\left.\begin{aligned}(\tau_{\text{coll}}^{-1})_{2p_{1/2}} &= nv\,\frac{8\pi e^4}{\hbar^2 v^2}\,\frac{3a_H^2}{4}\left[2ln\,\frac{mv^2}{2\hbar\omega_0} + ln\,\frac{mv^2}{10\hbar\omega_0} + ln\,\frac{mv^2}{6\hbar\omega_0}\right] \\ &= 24\pi\,\frac{n\hbar^2}{vm^2}\left[ln\left(\frac{mv^2}{2\hbar\omega_0}\right) - \tfrac{1}{4}ln\ 15\right] \\ &= \frac{1.5\times 10^{-4}}{\sqrt{T}}[ln(2.1\times 10^3 T) - 0.67]n\,.\end{aligned}\right\} (S')$$

The last form is convenient for numerical substitution, while the first is intended to show the relation of formula (S') to formula (S). The $|x_{sp}|^2$ would be replaced by $3a_H^2$ if there were no hyperfine structure. The coefficients $\frac{2}{4}$, $\frac{1}{4}$, $\frac{1}{4}$ are the relative strengths of transitions from $f = 1$ to $f = 1$, from $f = 0$ to $f = 1$, and from $f = 1$ to $f = 0$. The energy denominators $2\hbar\omega_0$, $10\hbar\omega_0$, $6\hbar\omega_0$ are the corresponding energy differences between s and p components. The frequency $\omega_0/2\pi$ corresponds to $\nu_0/c = 0.00049$ cm^{-1} of Figure 1.

For transitions from $2s_{1/2}$ to $2p_{3/2}$ by neglecting hyperfine structure one obtains

$$\left.\begin{aligned}(\tau_{\text{coll}}^{-1})_{2p_{3/2}} &= 48\pi \frac{n\hbar^2}{vm^2} ln \left(\frac{mv^2}{\hbar\omega(sp_{3/2})}\right) \\ &= \frac{3.0 \times 10^{-4}}{\sqrt{T}} nln(5.7T) .\end{aligned}\right\} \qquad (S'')$$

For $n = 30$ cm^{-3}, $T = 10{,}000°$ one has

$$(\tau_{\text{coll}}^{-1})_{2p_{1/2}} = \frac{1}{1400 \text{ sec}} , \qquad (\tau_{\text{coll}}^{-1})_{2p_{3/2}} = \frac{1}{1000 \text{ sec}} ,$$

and

$$\tau_{\text{coll}}^{-1} = (\tau_{\text{coll}}^{-1})_{2p_{1/2}} + (\tau_{\text{coll}}^{-1})_{2p_{3/2}} = \frac{1}{600 \text{ sec}} .$$

The value of the static effects corresponding to the foregoing n and T is $\tau^{-1} \sim 1/5000$ sec. For higher n the formulas (5) and (S) indicate an increased relative importance of the static effects. As has already been discussed in the beginning of this section, however, the static effects cease to have a definite meaning at higher electron densities, because at the smaller average distance of the electron from the atom, the electron distance changes by a large fraction of its value in a time small compared with the mean life of the 2p level.

It is clear that the foregoing estimates could be improved in many ways regarding the chance of transfer in a single collision, the effect of radiation before the transfer is completely effected, as well as with respect to the statistical nature of the field and the velocity distribution of electrons. Since at the density and temperature of electrons supposed to exist in interstellar space the effects of elec-

trons appear to be of secondary importance, such improved estimates do not appear to be essential at the present time.

II. SIMULTANEOUS EMISSION OF TWO PHOTONS

The chance per second for the simultaneous emission of two photons with one photon in the frequency range $d\nu'$ is

$$A\,d\nu' = \frac{1024\pi^6 e^4 \nu'^3 \nu''^3}{h^2 c^6}\left(\left|\sum_{n''}\left[\frac{(\boldsymbol{ru}')_{n'n''}(\boldsymbol{ru}'')_{n''n}}{\nu_{n''n}+\nu''} + \frac{(\boldsymbol{ru}'')_{n'n''}(\boldsymbol{ru}')_{n''n}}{\nu_{n''n}+\nu'}\right]\right|^2\right)_{\text{av}} d\nu' . \tag{6}$$

Here ν'' is the frequency of the second photon emitted simultaneously with ν'. Conservation of energy holds, so that

$$\nu' + \nu'' = \nu_{nn'} = \frac{E_n - E_{n'}}{h} .$$

The upper level (2s) is designated by n, the lower (1s) by n'. The intermediate levels are denoted by n''. They are in this case the p levels. Matrix elements are denoted by suffixes such as $(\boldsymbol{ru}')_{n'n''}$. The polarization vectors $\boldsymbol{u}'$, $\boldsymbol{u}''$ have absolute value 1 and are parallel to the electric intensity of the photons $h\nu'$, $h\nu''$. The displacement vector of the electron is denoted by $\boldsymbol{r}$. The averaging in the foregoing formula is supposed to be made over the directions of propagation and over the polarization directions independently for $h\nu'$ and $h\nu''$. The formula is supposed to be used for the probability of simultaneous emission of a pair of photons, and there is no distinction with respect to the order of emission. The total probability of emission is thus

$$\frac{1}{\tau} = A_t = \tfrac{1}{2}\int_0^{\nu_{nn'}} A\,d\nu' . \tag{6.1}$$

Equation (6) applies only if the electric dipole transitions to intermediate states predominate over other ways of obtaining these transitions, such as quadrupole and magnetic dipole effects. It applies in the case of the 2s state of hydrogen on account of the predominance of dipole transitions to the p levels. In applying equation (6), one may neglect the effects of hyperfine structure and of fine

structure since the operator $\mathbf{r}$ is diagonal in the nuclear and electronic spin. The s levels are then single, and every p level is triply degenerate. It is convenient to use for each energy three linearly independent p states having the transformation properties of the Cartesian co-ordinates x, y, and z. These eigenfunctions will be denoted by ξ, η, and ζ. The sum over n'' in equation (6) will be taken first over ξ, η and ζ for each energy $E_{n''}$. One has

$$\sum_{\xi,\eta,\zeta}(\mathbf{r}\mathbf{u}')_{n'n''}(\mathbf{r}\mathbf{u}'')_{n''n} = (xu_x')_{n'\xi}(xu_x'')_{\xi n} + (yu_y')_{n'\eta}(yu_y'')_{\eta n} + (zu_z')_{n'\zeta}(zu_z'')_{\zeta n}$$
$$= z_{n'\zeta}z_{\zeta n}(u_x'u_x'' + u_y'u_y'' + u_z'u_z'')$$

because

$$x_{n'\xi}x_{\xi n''} = y_{n'\eta}y_{\eta n''} = z_{n'\zeta}z_{\zeta n''}\ ; \qquad y_{n'\xi} = y_{n'\zeta} = \ldots = 0$$

by symmetry.

The sum over n'' is, therefore,

$$(\mathbf{u}'\mathbf{u}'')\sum_{m=2}\left[\frac{1}{\nu_{mn}+\nu''} + \frac{1}{\nu_{mn}+\nu'}\right] z_{n'm}z_{mn}\ .$$

Here the principal quantum number of the p state is denoted by m and the matrix elements $z_{n'm}$ are between the state n' and the state of type ζ with principal quantum number m. The factor $(\mathbf{u}'\mathbf{u}'')$ shows that there is a statistical relationship between the directions of polarization of the two photons and that the emission probability is proportional to the square of the cosine of the angle between the polarization vectors. Such a relationship is natural, since angular momentum must be conserved. The average over the polarization directions for $(\mathbf{u}'\mathbf{u}'')^2$ is $\frac{1}{3}$. Substituting into equation (6) one has

$$A\,d\nu' = \frac{1024\pi^6 e^4 \nu'^3 \nu''^3}{3h^2c^6}\left|\sum_{m=2}^{\infty} z_{n'm}z_{mn}\left(\frac{1}{\nu_{mn}+\nu'} + \frac{1}{\nu_{mn}+\nu''}\right)\right|^2 d\nu'\ . \quad (6.2)$$

The summation over the principal quantum number must be extended also into the continuous spectrum. The integral over the continuous spectrum is not indicated in the foregoing formula but is understood to be present.

In the evaluation of the sum one is helped by the summary of values of the squares of radial integrals given by Bethe[7] for the discrete spectrum and by the closed formulas of Gordon[8] and Stobbe[9] for the continuum. By using the Bohr radius as the unit of length, the wave functions will be used as follows:

$$\left.\begin{aligned} &\psi_{ms} = (4\pi)^{-1/2} R_{ms}\,, \qquad \psi_{mp} = \left(\frac{3}{4\pi}\right)^{1/2} \frac{z}{r} R_{mp}\,, \\ &\int_0^\infty r^2 R^2 dr = 1\,, \\ &R_{1s} = 2e^{-r}\,, \qquad R_{2s} = 2^{-1/2}\left(1 - \frac{r}{2}\right) e^{-r/2}\,, \\ &R_{2p} = 24^{-1/2} r e^{-r/2}\,, \qquad R_{3p} = \left(\tfrac{8}{27}\right) 6^{-1/2} e^{-r/3} r \left(1 - \frac{r}{6}\right). \end{aligned}\right\} \tag{6.3}$$

The notation

$$R_{mp}^{ns} = \int_0^\infty r^3 R_{ns} R_{mp} dr \tag{6.4}$$

will be used. One has

$$(m > 1) \qquad R_{mp}^{1s} = \left[\frac{2^8 (m-1)^{2m-5} m^7}{(m+1)^{2m+5}}\right]^{1/2}, \tag{6.5}$$

$$(m > 2) \qquad R_{mp}^{2s} = \left[\frac{2^{17} m^7 (m^2-1)(m-2)^{2m-6}}{(m+2)^{2m+6}}\right]^{1/2}, \tag{6.6}$$

and

$$R_{2p}^{2s} = -3\sqrt{3}\,, \tag{6.7}$$

where positive values of the square roots are understood. The radial integrals in Stobbe's notation are

$$C_{0,0}^{m_{r_1}} = C_{1s} = 4^2 \frac{e^{-(2/x)\arctan x}}{(1 - e^{-2\pi/x})^{1/2}} \frac{x^{1/2}}{(1+x^2)^{5/2}} \tag{6.8}$$

and

$$C_{1,0}^{m_{r_1}} = C_{2s} = 2^{17/2} \frac{e^{-(2/x)\arctan 2x}}{(1 - e^{-2\pi/x})^{1/2}} \frac{x^{1/2}(1+x^2)^{1/2}}{(1+4x^2)^3}\,, \tag{6.9}$$

[7] *Op. cit.*, Table 15, p. 442. [8] *Ann. d. Phys.*, **2**, 1031, 1929.

[9] *Ann. d. Phys.*, **7**, 661, 1930 (see Tables 2 and 3).

where $x = \kappa a_H$, $a_H = \hbar^2/me^2$, $\kappa = (2mE)^{1/2}/\hbar$. The superscript m_r is the radial quantum number for the continuum. To obtain the contribution to $\sum_m R^{1s}_{mp} R^{2s}_{mp}$ due to the continuum with a_H as unit of length one takes

$$\int C_{1s} C_{2s} dx \,.$$

The quantity x used here is Stobbe's x for the 1s state.

In terms of x the energy in the continuum is

$$E = x^2 \frac{e^2}{2a_H} \,.$$

Expressing the energy in Rydberg units $e^2/2a_H$, the energy is x^2. The smooth joining of the discrete and the continuous is verified, giving at the series limit $2^7 e^{-4}$ as the contribution per unit energy range with $e^2/2a_H$ as the unit of energy to $\sum_m (R^{1s}_{mp})^2$ and $2^{16} e^{-8}$ to $\sum_m (R^{2s}_{mp})^2$.

The relative signs of the radial integrals are important for the evaluation of equation (6.2). These can be checked by working out in terms of them

$$(r^2)_{1s,\,2s} = \int_0^\infty r^4 R_{1s} R_{2s} dr = -\tfrac{5}{2}\tfrac{1}{4}\tfrac{3}{3}\sqrt{2} = -2.98 \,,$$

which should also be expressible as

$$(r^2)_{1s,\,2s} = \sum_{m=2}^{\infty} R^{1s}_{mp} R^{2s}_{mp} + \int_{x=0}^{\infty} C_{1s} C_{2s} dx \,.$$

Using Bethe's table on his page 442 and estimating the contribution from $m = 8$ to $m = \infty$ as 0.10, one obtains $\sum_{m=3}^{\infty} R^{1s}_{mp} R^{2s}_{mp} = 2.38$ and $R^{1s}_{2p} R^{2s}_{2p} = -6.70$. The integral over the continuum is evaluated using Stobbe's numerical tables[9] of C_{1s}, C_{2s} as $+1.33$. This gives $(r^2)_{1s,2s} = 2.38 - 6.70 + 1.33 = -2.99$. The good check with the directly computed value indicates that no important slips have been made in the values of the matrix elements.

The radial integrals can be used also for the evaluation of the

products $z_{n'm}z_{mn} = \frac{1}{3}R^{1s}_{mp}R^{2s}_{mp}$ occurring in equation (6.2). The most significant contribution comes from $m = 2$. This contribution is negative, while the other contributions are positive. The sum, including the continuous spectrum, is negative. It will be convenient, therefore, to consider

$$-3\left(\sum_{m=2}^{\infty} + \int\right) z_{n'm}z_{mn}\left(\frac{1}{\nu_{mn} + \nu'} + \frac{1}{\nu_{mn} + \nu''}\right) = C\,.$$

The quantity C is positive, and the contribution to C due to $m = 2$ is positive, while the other contributions are negative. Expressing radial integrals in units a_H one has

$$C = \frac{8a_H^3 h}{3e^2}\left[-\sum_{m=2}^{\infty} R^{1s}_{mp}R^{2s}_{mp}\left(\frac{1}{\frac{1}{3} - \frac{4}{3}m^{-2} + y} + \frac{1}{\frac{4}{3} - \frac{4}{3}m^{-2} - y}\right)\right.$$
$$\left. - \int_0^{\infty} C_{1s}C_{2s}\left(\frac{1}{\frac{1}{3} + \frac{4x^2}{3} + y} + \frac{1}{\frac{4}{3} + \frac{4x^2}{3} - y}\right)dx\right],$$

where

$$\nu' = \nu_{nn'}y = \frac{3e^2 y}{8a_H h}\,.$$

The quantities in parentheses have a minimum for $y = \frac{1}{2}$ and increase steadily toward $y = 0$ and $y = 1$. Keeping the term for $m = 2$ as it stands and letting $y = \frac{1}{2}$ and 0 in the other terms, one obtains, therefore, as upper and lower limits for C,

$$C_2 \leq C \leq C_1\,,$$

$$C_1 = \frac{2a_H^3 h}{e^2}\left[\frac{a_1}{y(1-y)} - b_1\right], \qquad C_2 = \frac{2a_H^3 h}{e^2}\left[\frac{a_2}{y(1-y)} - b_2\right],$$

$$a_1 = a_2 = -\tfrac{4}{3}R^{1s}_{2p}R^{2s}_{2p}\,,$$

$$b_1 = \sum_{m=3}^{\infty} \frac{2R^{1}_{mp}R^{2s}_{mp}}{\frac{5}{8} - \frac{1}{m^2}} + \int_0^{\infty} \frac{2C_{1s}C_{2s}}{\frac{5}{8} + x^2}\,dx\,,$$

and

$$b_2 = \sum_{m=3}^{\infty} R_{mp}^{1s} R_{mp}^{2s} \left(\frac{1}{\frac{1}{4} - m^{-2}} + \frac{1}{1 - m^{-2}} \right) + \int_0^{\infty} C_{1s} C_{2s} \left(\frac{1}{\frac{1}{4} + x^2} + \frac{1}{1 + x^2} \right) dx .$$

Substituting upper and lower limits of C into equation (6.2) and then into equation (6.1), one obtains as upper and lower limits for $1/\tau$

$$\frac{1}{\tau} = \frac{3^4}{8^5} \left(\frac{e^2}{\hbar c} \right)^6 \left(\frac{e^2}{2a_H h} \right) \left[\frac{a_i^2}{6} - \frac{a_i b_i}{15} + \frac{b_i^2}{140} \right]$$

$$= 1.22 \left[\frac{a_i^2}{6} - \frac{a_i b_i}{15} + \frac{b_i^2}{140} \right] \text{sec}^{-1}$$

the upper limit corresponding to a_1, b_1 and the lower to a_2, b_2. The number $a_1 = a_2$ can be calculated exactly and is $(4/3)6.70 = 8.93$. The evaluation of b_1, b_2 has been made only approximately by using the radial integrals up to $m = 5$, neglecting $1/m^2$ for the remaining contributions of the discrete spectrum, and using $x = 0.4$ in the denominators of the integrals over the continuum. This gave $b_1 = 12.2$, $b_2 = 22.2$, and

$$4.4 \text{ sec}^{-1} < \frac{1}{\tau} < 8.7 \text{ sec}^{-1} .$$

The values of the denominators involving frequencies used for b_1 correspond to the maximum of the statistical weight factor $\nu'^3 \nu''^3$ at $\nu' = 3e^2/16a_H h$. It is, therefore, likely that the upper limit for $1/\tau$ is closer to $1/\tau$ than is the lower limit. It is probable, therefore, that

$$6.5 \text{ sec}^{-1} < \frac{1}{\tau} < 8.7 \text{ sec}^{-1} .$$

III. DOUBLE EMISSION FOR HELIUM AND OTHER EFFECTS

In *He* the $(1s\ 2s)\ {}^1S_0$ and $(1s\ 2s)\ {}^3S_1$ levels are also metastable. Direct transitions with emission of a single photon from $(1s\ 2s)\ {}^1S_0$ to the ground level $(1s)^2\ {}^1S$ are strictly forbidden for dipole or multipole radiation of either the electric or the magnetic type. Nevertheless,

double emission from (1s 2s) ^1S to the ground level should take place with roughly the same probability as from the 2s state in hydrogen. For the ^3S state, however, the spontaneous emission of two photons is very much smaller, since the intermediate states either combine poorly with ^3S or with ^1S. The factor due to this is of the order of magnitude of the admixture of $^3P_1$ in $^1P_1$, which is 1 $cm^{-1}/2000$ $cm^{-1} \sim 10^{-3}$ for matrix elements or $\sim 10^{-6}$ for the intensities. Double emission for ^3S should be, therefore, much less probable than for ^1S.

While single photon emissions are strictly forbidden from (1s, 2s) $^1S_0$ so that collisions and double emissions are the only effects, there are minor radiation effects in $^3S_1$ which compete with double emission. Without any attempt at precision the general order of expected magnitude of these effects will be considered.

The ordinary spin-orbit interaction is of the form

$$\boldsymbol{A}_1\boldsymbol{\sigma}_1 + \boldsymbol{A}_2\boldsymbol{\sigma}_2 ,$$

where $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$ are the Pauli spin vectors for the two electrons, and $\boldsymbol{A}_1$, $\boldsymbol{A}_2$ are orbital vectors. According to standard-selection rules, this interaction mixes $^1S_0$ only with $^3P_0$, and $^3S_1$ is mixed only with $^1P_1$ and $^3P_1$. The terms under consideration are even and, therefore, the P terms when expanded in terms of product wave functions are also even. They cannot be, therefore, of the (1s, np) or (2s, np) types, and it suffices to consider admixtures of pp′ and dd′ configurations. Configurations p^2 contribute only $^3P_0$ and $^3P_1$ but no $^1P_1$. Taking 1 cm^{-1} as the order of magnitude of the nondiagonal perturbation matrix element and using $\sim 300{,}000$ cm^{-1} for the distance between $(1s)^2$ and $(2p)^2$ and $\sim 150{,}000$ cm^{-1} for the distance between 1s 2s and $(2p)^2$, the coefficient of the $^3P_0$ wave function can be expected to be 10^{-5} or less for $(1s)^2$ as well as for 1s 2s. The admixture of $^1P_1$ in (1s 2s) $^3S_1$ is not likely to be bigger and will be considered as being taken care of by a coefficient 10^{-5} in the wave function.

In addition to the ordinary spin-orbit interaction there is present the spin-orbit-spin interaction of the type $T_2 = \frac{1}{2}[3(\boldsymbol{r}\boldsymbol{\sigma}_1)(\boldsymbol{r}\boldsymbol{\sigma}_2) - r^2(\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2)]$. It produces no mixing of $^1S_0$ since T_2 applied to a singlet wave function gives zero. In $^3S_1$ it produces an admixture of $^3D_1$ in

addition to $^3P_1$. The character of the terms is thus $^1S_0 + \alpha\ ^3P_0$ and $^3S_1 + \beta\ ^1P_1 + \gamma\ ^3P_1 + \delta\ ^3D_1$ where α, β, γ, and δ are $\sim 10^{-5}$, and the spectroscopic symbols stand for normalized wave functions.

In the foregoing estimates the $(2p)^2$ configuration was used in estimating the order of magnitude of the coefficients. The important coefficients will be seen to be α for $(1s)^2$ and γ.

The presence of terms in the pp′ configurations is suggested not only by the selection rules but also by the approximate form of the wave function. The orbital function for $^1S_0$ has the form $U_S(r_1, r_2; \mathbf{r}_1\mathbf{r}_2)$ where $\mathbf{r}_1, \mathbf{r}_2$ are, respectively, the vector distances from the nucleus to the electrons 1 and 2. The interaction energy $\boldsymbol{A}_1\boldsymbol{\sigma}_1 + \boldsymbol{A}_2\boldsymbol{\sigma}_2$ operating on the electronic wave function gives a linear combination of triplet spin functions and components of $(\boldsymbol{A}_1 - \boldsymbol{A}_2)U_S$. The vector $\boldsymbol{A}_1$ consists of two parts

$$\boldsymbol{A}_1 = \frac{\hbar e^2}{4m^2c^2}\{r_1^{-3}[\mathbf{r}_1 \times \mathbf{p}_1] + r^{-3}[(\mathbf{r}_1 - \mathbf{r}_2) \times (2\mathbf{p}_2 - \mathbf{p}_1)]\}\ .$$

The first part contains the angular momentum $\boldsymbol{L}_1 = [\mathbf{r}_1 \times \mathbf{p}_1]$. Because of it the contribution of $\boldsymbol{A}_1 - \boldsymbol{A}_2$ occurs only on account of the dependence of U_S on $\mathbf{r}_1 \cdot \mathbf{r}_2$ and not on r_1 or r_2. The orbital functions arising in this way are of the type $[\mathbf{r}_1 \times \mathbf{r}_2]\partial U_S/\partial(\mathbf{r}_1\mathbf{r}_2)$. The second part of $\boldsymbol{A}_1$ gives rise to orbital functions of the type

$$r^{-3}[\mathbf{r}_1 \times \mathbf{r}_2]\left\{\frac{\partial}{r_1\partial r_1} + \frac{\partial}{r_2\partial r_2} + \frac{2\partial}{\partial(\mathbf{r}_1 \cdot \mathbf{r}_2)}\right\} U_S\ .$$

If U_S did not depend on $\mathbf{r}_1\mathbf{r}_2$, only the second part of $\boldsymbol{A}_1$ would matter, and the admixture would be entirely of the pp′ type. This is in agreement with the fact that if the dependence of U_S on $\mathbf{r}_1\mathbf{r}_2$ is neglected, the function U_S is entirely of the ss' type so that d states of single electrons cannot be brought in by the spin-orbit interaction. If the dependence of U_S on $\mathbf{r}_1\mathbf{r}_2$ is taken into account, the pp′ configurations are not the only ones entering into the admixture.

Quadrupole radiation between $^3S_1 + \beta\ ^1P_1 + \ldots$ and $^1S_0 + \alpha\ ^3P_0$ is excluded by the exact selection rule which prohibits direct jumps between any two levels with angular momenta j', j'', for which a triangle with sides $j', j'', 2$ cannot be constructed.

The magnetic dipole operator consists of a part proportional to the total angular momentum which does not give rise to transitions. The remaining part of the magnetic dipole contains $\boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2$ and is proportional to the total electron spin. This part gives zero when applied to singlets and leaves the space function unchanged when

applied to other terms. The multiplicity of a term is also unchanged by $\boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2$. Hence the effect of the magnetic dipole is to give a transition only between $\gamma\ ^3P_1$ which is admixed to $^3S_1$ and $\alpha\ ^3P_0$ which is admixed to $^1S_0$. Using the estimate 10^{-5} for α and γ, one obtains the very small value

$$\left[\frac{10^{-5} \times 10^{-5}\hbar}{a_H mc}\right]^2 \sim 10^{-24}$$

for the ratio of the transition probability to that expected for an f value (oscillator strength) 1 with electric dipole. This effect is so small that one does not need a more accurate estimate.

In the absence of external fields and nuclear spin the $2s_{1/2}$ term of hydrogen is pure. The electric quadrupole and magnetic dipole effects vanish in this approximation, the first because one deals with an $s \rightarrow s$ jump and the second because the 1s and 2s states are strictly orthogonal. On account of the proton spin there is a slight perturbation of $s_{1/2}$ and $d_{3/2}$. The matrix element responsible for this does not vanish because the coupling of the electron to the proton involves Dirac's $\boldsymbol{\alpha}$, which in the Pauli approximation is neither a pure orbital nor a pure spin vector. One may expect therefore $s_{1/2}(f = 1)$ to be weakly mixed with $d_{3/2}(f = 1)$, but $s_{1/2}(f = 0)$ cannot be affected by this perturbation since $d_{3/2}$ has no $f = 0$. Even if the life of $2s_{1/2}(f = 1)$ should be shortened by electric quadrupole radiation from the $3d_{3/2}(f = 1)$, the $f = 0$ part could still show absorption lines. This perturbation appears to have only theoretical interest and should besides be small since the closest d term occurs for $n = 3$.

The proton presumably does not act on the electron quite as a point charge and may have a finite volume. As long as its action may be represented by a fixed central field, the 1s and 2s functions are orthogonal, and the magnetic dipole effect vanishes. If, however, the proton field depends on the velocity of the electron, this is no longer true, and other types of radiation could conceivably occur if the proton and electron were more intimately coupled. This already takes place weakly for Dirac's electron equation.

The approximate Pauli representation of the electron spin gives no magnetic dipole radiation for $2s_{1/2}$ of hydrogen. A slightly different result is obtained by using Dirac's equation. The matrix ele-

ments of the components of $\frac{1}{2}[\mathbf{r} \times \boldsymbol{\alpha}]$ take the place of the matrix elements of $\mathbf{r}$ in the formula for the transition probability due to the electric dipole. The radial integral $\int r^3(f_{1s}g_{2s} + f_{2s}g_{1s})dr$ occurs as a common factor in the matrix elements. Here g is the "large" radial function[10] and is approximately equal to the nonrelativistic one. Approximately f is a constant multiple of dg/dr. In this approximation the foregoing radial integral vanishes as is seen by partial integration using also the approximate orthogonality of g_{1s} and g_{2s}. The radial integral is not exactly zero, however, but only small. Its order of magnitude is $(e^2/\hbar c)^2(\hbar/mc)$. The emission probability is

$$\frac{\pi}{81}\left(\frac{e^2}{\hbar c}\right)^9 \nu \sim 5 \times 10^{-6}\ \text{sec}^{-1},$$

which is negligible. The foregoing relativistic magnetic dipole effect takes account of all terms in the usual expansion of the factor $e^{i\mathbf{kr}}$ in the expression for the vector potential up to $1 + i\mathbf{kr}$. It is sensitive to the nuclear charge on account of the extra factor $(Ze^2/\hbar c)^4$ due to the difference between the Dirac and the Schroedinger radial functions.

The writers are grateful to Drs. Struve, Henyey, and Unsöld for stimulating discussions.

UNIVERSITY OF WISCONSIN
AND
THE GEORGE WASHINGTON UNIVERSITY

[10] P. A. M. Dirac, *The Principles of Quantum Mechanics*, p. 252, Oxford: Clarendon Press, 1930.

THE GENERALIZED THOMAS-FERMI METHOD AS APPLIED TO STARS

R. E. MARSHAK AND H. A. BETHE

ABSTRACT

A treatment is given of stars of intermediate density such as dense red dwarfs and subdwarfs, using the statistical method of Thomas-Fermi to calculate the free-electron density, the gas pressure, and the opacities. The Thomas-Fermi equation has been generalized to take account of the high temperature and a mixture of elements. By taking finite boundary conditions as for metals, it is then possible to integrate the star equations.

In this note we wish to indicate a method of integrating the equations of stellar equilibrium for stars of moderately high central density (of the order of 1000 gm/cm^3). The method could have been used to determine the temperature-density distribution in the interior of a white dwarf star in a more elegant fashion than has been done;[1] but the labor involved in doing so would have been needlessly great since strong degeneracy and, therefore, complete (pressure) ionization prevail throughout most of the star.[2] However, it seems that it would be very worth while to investigate the internal constitution of the dense red dwarfs and the recently discovered subdwarfs[3] on the basis of a "generalized" Thomas-Fermi model; the information thus derived should be more accurate than that given by any of the usual stellar models and would therefore throw light on the nature of the energy-production process of this interesting group of stars.

The starting-point, then, is the Thomas-Fermi equation for electrons, generalized to include the effect of high temperatures and subject to finite boundary conditions as for electrons in metals. Solutions of this equation lead directly to expressions for the free-electron density, the gas pressure due to electrons, and the screening radius which enters the formula for the conductive opacity. Moreover, the

[1] R. E. Marshak and H. A. Bethe (to be published shortly).

[2] Moreover, it would be necessary to take account of relativistic degeneracy; this introduces additional difficulties for the Thomas-Fermi method. In this note we restrict ourselves to the nonrelativistic Thomas-Fermi atom.

[3] G. P. Kuiper, *Ap. J.*, **89**, 548, 1939.

mass density can be expressed in terms of quantities which the generalized Thomas-Fermi equation yields, and the star equations can then be integrated. These considerations hold for one element but can easily be extended to a mixture of elements such as the "Russell" mixture.

We write for the electron density n

$$n = \frac{2(mkT)^{3/2}}{(2\pi)^{3/2}\hbar^3} \cdot \frac{4}{3\sqrt{\pi}} (\log \alpha')^{3/2} \left[1 + \frac{\pi^2}{8 \log^2 \alpha'} \right] .^4 \qquad (1)$$

Here $\hbar$ = Planck's constant divided by 2π, m is the mass of the electron, k is Boltzmann's constant, T is the temperature, and α' is related to the Fermi energy, ζ, and the potential energy in the neighborhood of the atom, eV, by the equation

$$\log \alpha' = \frac{\zeta + eV}{kT} = \frac{e}{kT} \left(\frac{\zeta}{e} + V \right) = \frac{eU}{kT} . \qquad (2)$$

Then equation (1) becomes

$$n = \frac{2^{3/2} m^{3/2} e^{3/2}}{3\pi^2 \hbar^3} U^{3/2} \left[1 + \frac{\pi^2 k^2 T^2}{8 e^2 U^2} \right] . \qquad (3)$$

This electron density when inserted into the Poisson equation $\Delta U = 4\pi ne$ leads to an equation of the Thomas-Fermi form

$$\frac{d^2\Phi}{dx^2} = \frac{\Phi^{3/2}}{x^{1/2}} \left[1 + \frac{\beta x^2 T^2}{\Phi^2} \right] . \qquad (4)$$

[4] In this approximation only the leading temperature-dependent term is retained; if all orders were included, the expression in the brackets would be replaced by

$$\left[1 + 2 \sum_{\nu = 2, 4, 6 \ldots}^{\infty} \{ (1 - 2^{1-\nu}) \xi_\nu (-\log \alpha')^{-\nu} \tfrac{3}{2} \cdot \tfrac{1}{2} \ldots . (\tfrac{5}{2} - \nu) \} \right] ,$$

where ξ_ν is the Riemann-zeta function of order ν (cf. Sommerfeld, *Zs. f. Phys.*, **47**, 9, 1928). Also the exchange term is neglected and justifiably so (cf. H. Jensen, *Zs. f. Phys.*, **111**, 373, 1938).

In equation (4) we have set[5]

$$U = \frac{Ze}{r}\Phi \qquad r = ax\,, \tag{4a}$$

$$a = \left(\frac{9\pi^2}{128}\right)^{1/3} a_0 Z^{-(1/3)}\,, \qquad a_0 = \frac{\hbar^2}{me^2} = \text{Bohr radius}\,, \tag{4b}$$

$$\beta = \frac{3^{4/3}\pi^{10/3}}{2^{29/3}Z^{8/3}}\left(\frac{k}{Ry}\right)^2, \qquad Ry = \frac{me^4}{2\hbar^2} = \text{Rydberg energy}\,. \tag{4c}$$

This equation is to be integrated subject to the boundary conditions

$$(A)\ \Phi(0) = 1 \qquad (B)\ \frac{dn}{dx} = 0 \quad \text{at} \quad x = X\,.$$

The first condition implies that in the neighborhood of the nucleus the potential goes as Ze/r, the second, that the atom occupies a spherical volume with radius $R = aX$. The second condition is equivalent to the requirement that Z electrons be contained in the sphere of radius R, i.e., $4\pi\int_0^R nr^2dr = Z$. This can be seen either by direct integration using $n = \Delta U/4\pi e$ or from the fact that the electric field must vanish at the surface of a neutral atom; this means that $dV/dx = 0$ at the boundary and, since for a given temperature V is a unique function of n, also $dn/dx = 0$. Equivalent to $dn/dx = 0$ is the condition[6] $d\Phi/dx = \Phi/x$. For a given X the boundary condition $dn/dx = 0$ at $x = X$ will require a certain value $c(X)$ for the derivative $d\Phi/dx$ at $x = 0$.

Once equation (4) is integrated for given temperature and mass density, the free-electron density can be found immediately from

$$n = \frac{32Z^2}{9\pi^3 a_0^3 X^{3/2}}\Phi_X^{3/2}\left[1 + \frac{\beta T^2 X^2}{\Phi_X^2}\right], \tag{5}$$

and the electron pressure from

$$p_G^{\text{elec}} = \frac{2^{25/3}Z^{10/3}e^2}{5\cdot 3^{8/3}\pi^{11/3}a_0^4}\frac{\Phi_X^{5/2}}{X^{5/2}}\left[1 + \frac{5\beta T^2 X^2}{\Phi_X^2}\right]. \tag{6}$$

[5] E. Fermi, *Zs. f. Phys.*, **48**, 74, 1926.

[6] J. C. Slater and H. Krutter, *Phys. Rev.*, **47**, 559, 1935.

Equation (6) follows from the definition of pressure as the mean rate of transfer of momentum across unit surface.[7] A solution of equation (4) would give the screening radius $R = aX$ directly, and this expression could be used to evaluate the conductive opacity (cf. eqs. [33] and [41] in n. 1). Finally, the mass density ρ is given by

$$\rho = \frac{3Am_H}{4\pi R^3}; \qquad A \text{ is atomic weight}. \tag{7}$$

The above formulae hold for one element; when a mixture of elements is present the only modification introduced is one of self-consistency: the gas pressure due to the electrons and, therefore, the free-electron density must be the same at the boundary of each atom.[8] The mass density is then given by

$$\rho = \frac{3m_H}{4\pi \sum_Z \frac{c_Z}{A_Z} R_Z^3} = \frac{3.89}{\sum_Z \frac{c_Z}{A_Z Z} X_Z^3}. \tag{7a}$$

In equation (7*a*) c_Z is the concentration of element Z.

We have shown how certain astrophysically significant quantities can be expressed in terms of solutions of equation (4). Since this equation can be integrated only by numerical means, we point out two ways of doing this. The first is straightforward and involves a numerical integration by a simple polynomial approximation.[9] The second is to treat the temperature-dependent term in equation (4) as a perturbation and to make use of tables[10] for the unperturbed Thomas-Fermi function which have already been calculated. In the latter case, if we let $\Phi = \Phi_0 + \Phi_1$, then Φ_1 satisfies the differential equation

$$\frac{d^2\Phi_1}{dx^2} - \frac{3\Phi_0^{1/2}}{2x^{1/2}\Phi_1} = \frac{\beta T^2 x^{3/2}}{\Phi_0^{1/2}}, \tag{8}$$

[7] $p = \frac{1}{3}\int_0^\infty N(q) \cdot q \cdot v_q dq$, where $N(q)dq$ denotes the number of electrons per unit volume with momenta between q and $q + dq$, and vq is the velocity associated with momentum q. The integral can be evaluated by means of Sommerfeld's formulae (cf. Chandrasekhar, *An Introduction to the Study of Stellar Structure*, 1939).

[8] This essentially fixes the X for each atom and replaces the condition obtained from arguments about the free atom (cf. n. 6).

[9] M. F. Manning and J. Millman, *Phys. Rev.*, **53**, 673, 1938.

[10] Professor Slater possesses many copies of these tables; we are indebted to him for sending us a copy.

with the boundary conditions

$$(A') \quad \Phi_1(0) = 0 \qquad (B') \quad \frac{(\Phi_1)_X}{X} = \left(\frac{d\Phi_1}{dx}\right)_X .$$

In this derivation account has been taken of the fact that

$$\frac{d^2\Phi_0}{dx^2} = \frac{\Phi_0^{3/2}}{x^{1/2}}$$

with

$$\Phi_0(0) = 1 , \qquad \left(\frac{d\Phi_0}{dx}\right)_X = \left(\frac{\Phi_0}{x}\right)_X .$$

In terms of Φ_0 and Φ_1 we have for the free-electron density

$$n = \frac{32Z^2}{9\pi^3 a_0^3}\left(\frac{\Phi_0}{x}\right)_X^{3/2} + \frac{16Z^2}{3\pi^3 a_0^3}\left(\frac{\Phi_0^{1/2}\Phi_1}{x^{3/2}}\right)_X + \frac{32Z^2\beta}{9\pi^3 a_0^3}\left(\frac{x}{\Phi_0}\right)_X^{1/2},$$

and for the pressure

$$p_G^{\text{elec}} = \frac{2^{25/3}Z^{10/3}}{5\cdot 3^{8/3}\pi^{11/3}}\frac{e^2}{a_0^4}\left(\frac{\Phi_0}{x}\right)_X^{5/2} + \frac{2^{47/6}Z^{10/3}}{3^{8/3}\pi^{11/3}}\cdot\frac{e^2}{a_0^4}\left(\frac{\Phi_0^{3/2}\Phi_1}{x^{5/2}}\right)_X^{5/2} + \frac{2^{25/3}Z^{10/3}}{3^{8/3}\pi^{11/3}}\frac{e^2\beta}{a_0^4}\cdot\left(\frac{\Phi_0}{x}\right)_X^{1/2}.$$

Some rough calculations we have made indicate that mass densities of the order of a thousand fall within the range of the Slater-Krutter tables. The Thomas-Fermi procedure is thus quite feasible for the dense red dwarfs and the subdwarfs and should reveal a great deal about the temperature-density distribution of these stars, provided the hydrogen contents are not too high.

UNIVERSITY OF ROCHESTER
AND
CORNELL UNIVERSITY

PROPER MOTIONS IN THE GALACTIC CLUSTER M 67

E. G. EBBIGHAUSEN

ABSTRACT

Proper motions are derived on Yerkes plates for an area 30′ in diameter centered on the cluster M 67. The results are contained in Table 2. Plots of the motions are given in Figures 1, *A* and 1, *B*, and the position of the cluster members in the cluster in Figures 2, *A* and 2, *B*. Figure 3 gives the spectrum-magnitude diagram of the cluster, and Figure 4 the luminosity function of the cluster members.

This paper gives the results of the determinations of proper motions in the galactic cluster M 67 ($\alpha = 8^h45^m8$; $\delta = +12°11'$; $l = 184°$; $b = +33°$). For this cluster Trumpler gives 2–3a as its type, meaning that the cluster has many giants and that the upper limit of the main sequence is at spectral type A. For M 67 he gives 740 parsecs as its distance and 16′ as its apparent diameter. Because of its large distance from the galactic plane (+406 parsecs) and the large number of late-type giants, the cluster is of considerable interest.

Previous work on the cluster has been done by Olsson at Stockholm in 1898[1] by Fagerholm,[2] Shapley,[3] and van Rhijn.[4] The results of these papers will be discussed later when the occasion demands.

I. THE MATERIAL

The material consisted of four pairs of plates, each pair having only one image of each star. Table 1 is a list of these plates. I wish

TABLE 1

Old Plate	New Plate	Interval in Years	Limiting pv Magnitudes	Symbol
F 179	F 599	26.1	13.0	P I
F 174	F 600 A	26.1	13.0	P II
F 305	F 597 A	17.0	12.3	P III
Sch 241	F 601 A	35.0	11.7	P IV

to express my thanks to Dr. J. Titus and Dr. C. Hetzler for having taken these plates in my absence.

[1] *Pub. Stockholm Obs.*, **6**, No. 4, 1898.

[2] Diss. Upsala, 1906.

[3] *Mt. W. Contr.*, No. 117, 1916.

[4] *Groningen Pub.*, No. 33, 1922.

II. MEASUREMENT AND REDUCTION

The procedure adopted in this paper is the same as that used in the author's paper[5] on the galactic cluster NGC 752, where it is fully described. The only difference in procedure is that the measurements were made only with the north edge upward and not again with the same edge to the right. This resulted in two measures of Δx and two of Δy, which were then averaged to form for each star a mean value $\overline{\Delta x}$ and $\overline{\Delta y}$. A least-squares solution using a linear formula was then made for each set of differences (Δx and Δy), using all the stars on P I. With the aid of the resulting constants, the motions μ_x and μ_y were derived. The stars of large motion were removed, and the differences ($\overline{\Delta x}$ and $\overline{\Delta y}$) of the remaining stars were used in a final solution of the reduction constants and in the derivation of the motions of the stars on P I. These comparison stars were used whenever possible on the remaining three plates.

The removal of the magnitude equation was accomplished in the following manner: For all stars common to all four plates the motions in the x- and y-directions were averaged and then plotted. From the resulting concentration near the origin, a number of stars closest to the center of gravity were taken as being cluster stars. Then for each plate the individual values of μ_x and μ_y were plotted against the magnitude, and through the points on each plot a straight line was drawn. For five of the resulting eight plots the line had an appreciable slope, and the average value of the magnitude equation so derived was about $0''.0025$/year mag. The magnitude equation for the stars on P I and P II fainter than 11.7 was obtained by extrapolating the straight line defined by the brighter stars.

After the motions from the four plates had been corrected, they were then in a form to permit them to be combined after the weights of the individual plates had been determined. In order to do this the differences P I − P II, P I − P III, P II − P III, P III − P IV were formed for all stars common to all four plates and for the motions in both x and y. Each of the eight groups of differences was divided into two subgroups according to whether the star was brighter than or fainter than $m_{pv} = 11.2$. Then, for each of the sixteen subgroups the quantity $\Sigma v^2/(n-1)$ was formed. For none of the sub-

[5] *Ap. J.*, **89**, 431, 1939.

TABLE 2

Star No.	m_{pv}	Spec.	μ_x	μ_y	Wt.	Membership
2	13.1	.....	+ 52	− 19	1	3
3	12.7	.....	−116	0	2	4
4	12.7	.....	− 21	+ 14	2	2
5	12.9	.....	+ 20	+ 15	1	2
6	12.7	.....	− 30	− 8	2	2
8	12.4	.....	− 28	+ 48	2	3
12	13.0	.....	+ 27	− 9	1	2
16	12.6	.....	− 19	− 2	2	2
18	12.8	.....	+ 14	+ 44	2	3
20	12.8	.....	− 12	+ 3	2	2
22	12.8	.....	+ 20	+ 30	2	2
23	12.4	.....	− 86	+145	2	4
24	13.0	.....	− 17	− 51	1	3
28	12.8	.....	− 10	− 2	2	2
29	13.2	.....	+ 70	− 36	1	3
30	12.0	F2	− 20	+ 3	3	1
34	13.0	.....	+ 37	− 15	1	2
36	13.2	.....	+ 68	− 35	1	3
37	12.7	.....	+ 39	+ 33	2	3
39	13.1	.....	+ 28	+ 29	1	2
41	13.1	.....	+ 50	+ 31	1	3
44	13.1	.....	− 22	− 28	2	2
45	13.0	.....	+ 45	− 61	2	3
46	13.0	.....	+ 32	+ 18	2	2
47	11.5	.....	+126	−198	3	4
48	12.7	.....	+ 41	− 6	2	2
49	12.6	.....	+ 96	+ 36	2	4
51	12.7	Go	− 20	− 17	2	2
53	13.2	.....	+ 38	− 4	1	2
54	12.7	.....	0	− 15	2	2
55	11.6	A4	+ 46	+ 7	4	3
56	13.2	.....	+ 99	− 13	1	4
65	12.8	F4	+ 16	− 29	2	2
68	13.2	.....	+129	+ 75	1	4
69	13.2	.....	− 67	− 34	1	3
70	11.7	F2	+102	+ 56	3	4
72	12.4	K:	− 24	+ 1	1	4
78	13.3	.....	+ 75	+ 20	1	3
78	13.6	.....	+104	+ 6	1	4
79	12.6	.....	− 12	− 46	2	3
81	10.2	B9	0	+ 18	4	1
83	13.0	.....	+ 31	− 29	1	2
84	10.6	gK1	+ 10	− 8	4	1
88	12.4	Go	− 4	− 12	2	2
90a	11.2	F3	+ 2	+ 25	4	1

TABLE 2—*Continued*

Star No.	m_{pv}	Spec.	μ_x	μ_y	Wt.	Membership
90β.....	12.5	.....	+ 1	− 10	1	2
91.....	12.7	F2:	+ 42	− 30	2	3
92a.....	12.3	.....	+ 14	+ 27	2	2
94.....	12.7	F8:	+ 24	+ 18	2	2
95.....	12.8	F2	0	+ 11	2	2
96.....	12.6	F8	+ 29	+ 3	2	2
98.....	12.7	F5	+ 3	− 3	2	2
101.....	12.9	.....	+ 17	+ 25	1	2
102.....	12.4	.....	− 16	− 17	2	2
104.....	11.4	gG5	+ 4	− 19	3	1
105.....	10.4	gKo	− 5	− 6	4	1
108.....	9.9	gK3	− 1	+ 8	4	1
111.....	12.6	.....	− 4	− 11	2	2
112.....	13.0	.....	+ 20	− 6	1	2
115.....	12.7	F6	− 1	− 4	2	2
116.....	12.9	.....	+ 10	− 22	1	2
117.....	12.6	.....	+ 22	+ 3	2	2
119.....	12.8	F8	0	+ 28	2	2
124.....	12.2	F1	− 2	+ 28	3	1
127.....	12.7	.....	− 10	− 25	2	2
128.....	13.0	.....	− 39	− 34	2	3
130.....	12.9	G:	+ 20	− 9	2	2
131.....	11.3	F1	+ 19	0	4	1
132.....	13.1	G:	− 16	− 27	2	2
134.....	12.3	F5	+ 22	− 1	3	1
135.....	11.5	gKo	+ 18	+ 36	3	2
136.....	11.4	F9	+ 18	− 4	4	1
141.....	10.4	gKo	− 13	+ 9	4	1
143.....	11.6	gG8	+ 7	− 10	4	1
144.....	12.9	.....	+ 82	− 2	2	3
145.....	12.8	Go:	− 12	− 10	2	2
147.....	13.0	Go:	+ 22	− 32	2	2
148.....	13.0	Go	+ 22	+ 20	2	2
149.....	12.7	F5	+ 16	+ 4	2	2
151.....	10.6	G9	+ 18	+ 4	4	1
153.....	11.5	A3	− 9	− 2	4	1
155.....	10.5	gGo	− 49	+ 7	4	3
156.....	11.2	A1	+ 16	− 12	4	1
157.....	12.7	F8	+ 14	− 4	2	2
160.....	12.8	.....	− 7	+158	2	4
161.....	12.8	F3	+ 10	+ 20	2	2
162.....	12.8	F8	+ 4	+ 32	2	2
163.....	12.7	Go	+ 6	− 31	1	2
164.....	10.8	gG9	+ 16	− 7	4	1
165.....	12.7	F5	+ 53	− 12	2	3

TABLE 2—*Continued*

Star No.	m_{pv}	Spec.	μ_x	μ_y	Wt.	Membership
166	12.8	G5	+ 40	− 24	2	2
170	9.8	gK2	− 7	+ 2	4	1
171	12.9	.....	+ 58	+ 10	1	3
172	12.6	.....	+ 6	− 36	2	2
173	12.2	F8	+ 28	− 3	3	1
173	13.0	.....	+ 36	− 4	2	2
174	12.7	F8	+ 39	− 6	2	2
175	13.4	G:	+ 31	− 4	1	2
176	12.7	F8	+ 25	− 22	2	2
180	12.6	F5	− 10	− 22	2	2
181	12.8	.....	+ 30	− 10	2	2
182	12.7	.....	+ 7	0	2	2
184	12.3	Fo	− 9	+ 21	3	1
185	11.3	A4	+ 9	+ 2	4	1
185	12.7	F6	+ 20	− 15	1	2
187	13.0	F8	− 17	+ 6	2	2
189	12.7	.....	+ 22	− 27	2	2
190	11.2	A8	+ 14	− 4	4	1
192	12.8	.....	+ 16	0	2	2
193	12.3	gG5	+ 34	− 6	2	2
195	12.7	Go	+ 42	+ 23	2	2
199	13.0	.....	+ 46	− 18	2	3
200	11.0	G8	−172	+ 54	4	4
202	12.7	F5	+ 20	− 36	2	2
203	12.9	.....	+ 18	+ 10	2	2
206	12.7	.....	−198	− 20	2	4
207	12.3	.....	+ 55	− 12	3	3
210	12.3	F8	+ 15	+ 20	3	1
215	12.7	F5	+ 20	+ 27	2	2
216	12.6	F5	+ 80	− 24	2	3
217	11.4	Ko	+ 22	+ 18	4	1
217	12.2	.....	− 4	+ 11	3	1
218	11.7	.....	+ 6	− 6	4	1
219	12.6	.....	+ 60	+ 28	2	3
221	12.6	F8	+ 68	− 88	2	4
223	10.7	gG8	+ 12	− 7	4	1
224	11.1	gK2	− 11	+ 4	3	1
225	12.9	.....	+ 10	− 26	2	2
226	12.7	Go	+ 1	− 22	2	2
227	12.7	Go	+ 27	+ 10	2	2
228	12.9	.....	− 9	− 54	1	3
229	12.8	F2	+156	−139	2	4
231	11.6	gKo	− 6	0	3	1
234	11.6	.....	− 27	−509	3	4
236	12.4	Go	+ 43	− 1	1	2

TABLE 2—*Continued*

Star No.	m_{pv}	Spec.	μ_x	μ_y	Wt.	Membership
237	12.7	.....	+ 2	+ 27	2	2
238	11.2	A3	+ 16	+ 6	4	1
241	12.7	.....	+ 28	− 6	2	2
242	8.0	gK0	− 60	+ 20	3	4
243	12.7	F8	+ 14	+ 2	2	2
244	11.0	G6	− 2	+ 3	4	1
248	12.3	F2	− 12	0	2	2
255	12.7	G0	+ 28	− 6	2	2
256	12.6	.....	+ 18	0	2	2
257	11.6	F8	−128	−366	4	4
261	10.7	A3	+ 12	− 78	4	4
262	12.7	.....	+ 59	+ 26	1	3
265	12.5	F8	− 27	+152	3	4
266	10.7	gG5	+ 13	− 24	4	1
267	12.7	.....	+ 16	+ 62	2	3
269	13.0	.....	+ 26	+ 16	1	2
271	12.8	.....	+ 33	− 22	2	2
277	13.0	.....	+ 74	+110	2	4
280	11.0	A2	+ 33	+ 34	2	2
281	12.9	.....	+ 30	− 6	1	2
285	12.5	.....	+210	− 86	2	4
286	10.7	gG5	+ 21	+ 50	1	3
289	12.7	.....	+ 33	+ 3	3	1

groups did this quantity differ significantly from the mean of all, and hence all plates were given unit weight regardless of co-ordinate or magnitude. On this basis a weight of four corresponds to a mean error of $\pm 0''.0012$/year in either co-ordinate.

With the aid of these weights the motions from the various plates were combined and are given in Table 2.

The columns give, respectively:

1. The number according to Fagerholm's catalogue.[2] His paper also gives a chart of the cluster.

2. The photovisual magnitude according to Shapley.[3] His paper also gives photographic magnitudes.

3. The spectral type, kindly communicated in advance of publication by Dr. R. J. Trumpler.

4 and 5. The proper motions in x and y expressed in units of $0''.0001$/year.

6. The weights of the motions in the preceding columns. Weight 4 corresponds to a mean error of $\pm 0''.0012$/year.

7. A measure of the probability that a star is a member of the cluster. Unity indicates that the star has a high probability of being a cluster member; 2 indicates that the probability is lower but that the star has a reasonable chance of being a member; 3 signifies that the star is probably not a member; and 4 denotes that it is definitely not a cluster member. The assignment of weights is discussed in section III.

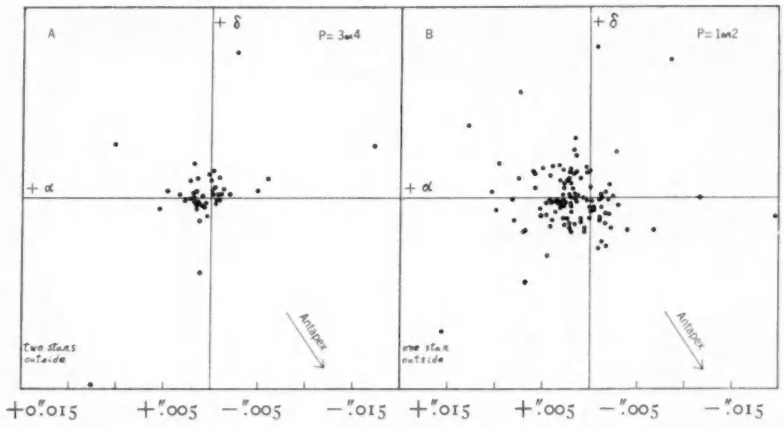


FIG. 1

III. DISCUSSION OF THE RESULTS

The data in Table 2 are complete down to $m_{pv} = 13.0$, within a radius of 15′ from star No. 155, which was considered by Shapley to be the center of the cluster.

Figure 1, *A*, is a plot of the motions of the stars with a weight of 3 or 4, and Figure 1, *B*, is a plot of those whose weights are either 1 or 2.

The assignment of probability of cluster membership was made in the following manner: The average weight of the stars in Figure 1, *A*, corresponds to a mean error of $\pm 0''.0014$/year. About the center of gravity of the concentration three circles were drawn with radii equal to 2, $2\sqrt{2}$, and 4 times the mean error just given. Stars within the smallest circle were given class 1 membership, those between the

first and second circles class 2, those between the second and third circles class 3, and those outside the largest circle class 4. The average weight of the stars in Figure 1, *B*, corresponds to a mean error of $\pm 0''.0021$/year. In this case the circles were given radii of two and four times the mean error just given, and no class 1 membership was assigned. Those stars within the first circle were given class 2 membership and those outside the second circle class 4.

Figure 2, *A*, gives the position in the cluster of all stars with weights 3 or 4, and Figure 2, *B*, gives the positions of all whose

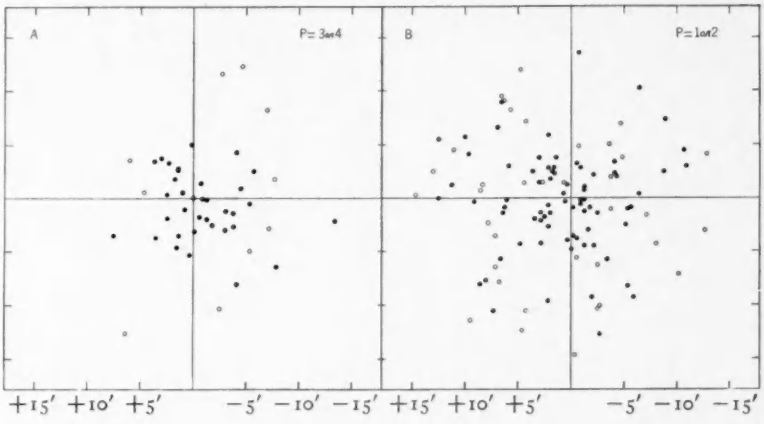


FIG. 2

weights are 1 or 2. In both cases the filled dots correspond to stars whose membership class is 1 or 2.

The appearance of the cluster on a photographic plate is that of a nucleus of bright stars about 16′ in diameter, superimposed on a background of fainter stars. On P III and P IV only the nucleus of brighter stars was present, while on P I and P II a considerable number of fainter stars outside the nucleus were measured also. For this reason, stars with weight 3 or 4 are concentrated toward the center of the measured area, and stars of weight 1 or 2 are spread over the whole measured area. On the basis of star counts Shapley has found that the cluster of bright stars is "merely a well marked nucleus of brighter and redder stars in a much larger system," which he estimates may extend to a radius of as much as half a degree from

the center. This is the "shoulder" effect which has been found for several clusters by Trumpler.

Figure 3 is the magnitude-spectrum diagram for the cluster. The ordinates are bolometric magnitudes, and the abscissas are given as both log *Te* and spectral type. The open circles refer to stars whose

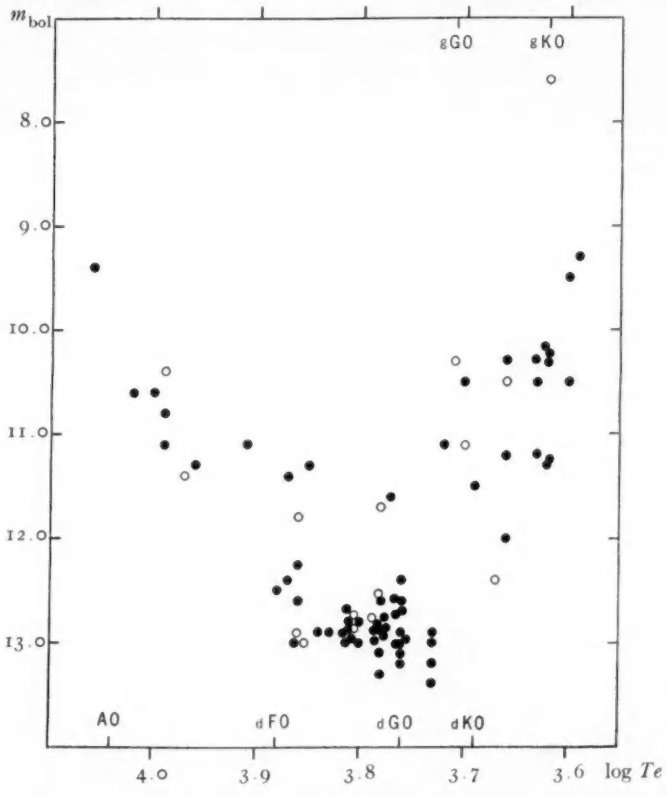


FIG. 3

membership class is 3 or 4. Attention is called to star No. 193, whose position in this diagram is log $Te = 3.66$ and $m_{bol} = 12$. The distance modulus being 9.3 mag., the star should have an $m_{bol} = +2.7$ which, if the star belongs to the cluster (membership class 2), makes it rather faint for its spectral type, gG5.

The brightest star in the cluster, No. 242, appears not to be a

member. However, since it is 2 mag. brighter than the next faintest star in the cluster, the corrections for magnitude equation had to be extrapolated, and therefore the corrections are very uncertain. The corrected values of motions as derived from the three plates on which the star was measured are quite discordant and their mean cannot be very accurate. Since the magnitude equation is uncertain, it might be better to use the uncorrected values. These happen to be very much more in accord with one another, but their mean still gives

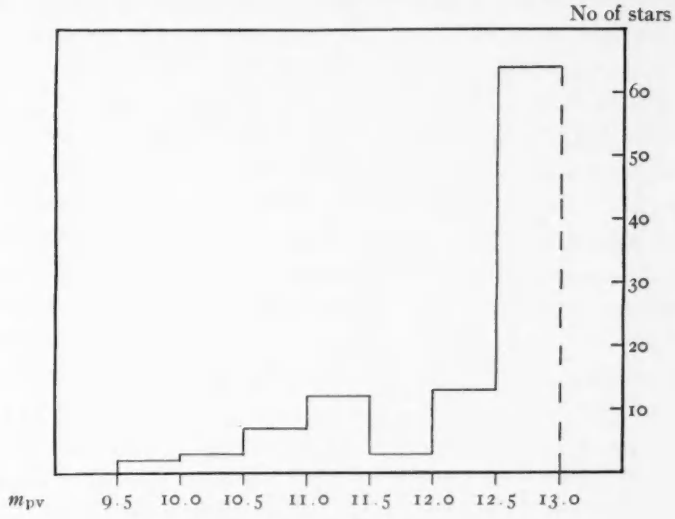


FIG. 4

the star a class membership of 3. The considerable deviation of this star from the center of the cluster (7′) might also be considered as evidence that it does not belong to the cluster. The meridian proper motion of this star as derived from only two observations is $\mu_x = -0''.037$/year and $\mu_y = -0''.006$/year.

The spectrum-magnitude diagram of M 67 resembles, to a considerable extent, that obtained for globular clusters, in that a prominent late-type giant sequence is present.

Since the late-type giants have a smaller concentration toward the galactic plane than the B- and O-type stars, those clusters which contain late-type giants might be expected to behave in the same fashion

with respect to clusters that have no such giants but whose brightest stars are of types O, B, and A. To examine this expectation, the galactic clusters in Trumpler's[6] list were divided into three groups, as follows: first, those whose Trumpler type is 1, i.e., no giants; second, those whose type is 1–2, i.e., uncertainty exists as to whether the giants belong or not; and, third, 2, 2–3, or 3, i.e., giants are certainly members of the cluster. For each of these groups using Trumpler's values for the z-distances the arithmetic mean was formed, and these values together with their standard deviation are for the three groups, respectively, 48 ± 15, 60 ± 16, and 102 ± 32 parsecs. In order to avoid selection effects, only clusters between 200 and 1000 parsecs were used in the computation. The run of decreasing concentration toward the galactic plane is in the direction to be expected for early-type and late-type stars not belonging to clusters.

The luminosity function is given in Figure 4. The rapid increase of stars beginning at $m_{pv} = 12.0$ may, to a small extent, be due to the fact that the average motion of the cluster stars differs but little from the mean of the noncluster stars, and some noncluster stars may be included as members. This is partly offset by the fact that some stars which were given class 3 membership may actually belong to class 2.

IV. THE PROPER MOTION OF THE CLUSTER

Two previous attempts have been made to determine the proper motion of the cluster; one by Fagerholm[2] and the other by van Rhijn.[4] Fagerholm compared the positions of 118 stars common to his catalogue and that of Olsson's,[1] made eight years before, but found no distinct evidence of motion.

From two *carte du ciel* plates of the cluster with an interval of 27 years, van Rhijn determined the absolute proper motion of the cluster to be $0''.0073$/year in position angle 287°. The mean error of an individual star's motion in either co-ordinate is $\pm 0''.009$/year. Although this direction differs by 13° from the direction of the convergent for the Hyades (274°), van Rhijn thought that the two clusters might have the same motion in space. None of the stars measured in the region of M 67 is found in the *Boss General Catalogue*, and only

[6] *Lick Obs. Bull.*, **14**, 154, 1930.

one star, No. 242, is in Schorr's *Eigenbewegungs Lexikon* as mentioned before.

An attempt was made to determine the proper motion from the measures of faint stars in the cluster. The motion of the cluster relative to 29 faint noncluster stars whose mean photographic magnitude is 13.6 was computed to be $\mu_x = -0''.0017 \pm 0''.0014$/year and $\mu_y = -0''.0008 \pm 0''.0012$/year (m.e.). For stars of $m_{pg} = 13.6$ at galactic latitude $33^\circ.3$, Oort gives a mean parallax of $0''.0021 \pm 0''.0002$ (m.e.), where, because of the few stars used in our case, the mean error is a lower limit. The antapex is 58° from the cluster, in position angle 214°. On the basis of a solar motion of 20 km/sec, the parallactic motion of the group of faint stars used is $0''.0075 \pm 0''.0008$/year in position angle 214°. Hence, the proper motion of the cluster so determined is $0''.009 \pm 0''.002$/year in position angle $220^\circ \pm 10^\circ$ (m.e.). On the basis of this determination, the difference in direction between the motion of M 67 and the direction of the Hyades convergent is 54°, and hence van Rhijn's suggestion seems improbable.

It is of interest to reinvestigate the space motion of the Praesepe cluster in order to determine whether or not it is parallel to that of the Hyades. Of the stars listed by Klein Wassink[7] as being members of the Praesepe cluster, 33 were found to have proper motions in the *General Catalogue*. The weighted mean proper motion is $\mu_x = -0''.034 \pm 0''.001$/year and $\mu_y = -0''.018 \pm 0''.001$/year or a total proper motion of $0''.038 \pm 0''.001$/year in $242^\circ.1 \pm 1^\circ.5$. In Moore's catalogue of radial velocities 6 stars were found. Two of these are binaries; for one there is no orbit, and for the other a determination has been made, and hence its space motion is known. The weighted mean of the five determinations of radial velocity is $+33.2 \pm 1.3$ km/sec. The parallax of the cluster as derived by Trumpler is $0''.0067$. On the basis of these data the calculated position of the convergent is $\alpha = 95^\circ.2$ and $\delta = -0^\circ.6$, with a mean error of about one degree in either co-ordinate. The computed space velocity is 42.9 ± 1.5 km/sec. The distance between this convergent and that of the Hyades as determined by Smart[8] is $9^\circ.5$. An attempt was then made to reduce this distance by changing the adopted parallax of Praesepe.

[7] *Groningen Pub.*, **41**, 1927. [8] *M.N.*, **99**, 3, 1939.

By changing the parallax to $0''.0072$, the distance between the convergents was reduced to $9^{\circ}.3$, but no more. Hence it would seem improbable that the magnitude and direction of the space motion of the two clusters is the same.

I wish to express my thanks to Dr. Struve for having placed the facilities of the Yerkes Observatory at my disposal during the summer of 1939 and gratefully to acknowledge the advice which Dr. Kuiper has given to me.

YERKES OBSERVATORY
September 1939

THE HYPOTHESIS OF THE EXISTENCE OF CONTRATERRENE MATTER

V. ROJANSKY

ABSTRACT

Symmetry arguments suggest the possibility of existence of atoms (here called "contraterrene") consisting of negatively charged nuclei surrounded by positrons. In this note it is assumed that a fundamental contraterrene particle and the corresponding ordinary (here called "terrene") particle annihilate each other on collision. Annihilation effects should make it possible to identify contraterrene matter (which presumably cannot be distinguished from terrene matter spectroscopically) provided that it can be observed in the process of colliding with terrene matter. Within the solar system perhaps the only bodies that are not certainly terrene are some meteors and the comets.

The quantum theory of the electron suggests a symmetry between positive and negative electric charges, and it has been conjectured that in some parts of space there might exist atoms whose nuclei are charged negatively and whose extranuclear members are positrons.[1] The purpose of this note is first to put the speculative notion of such atoms into a reasonably clear-cut form and then to consider briefly some of the astrophysical implications that their existence might have. For brevity, we shall call these atoms and their constituents and aggregations "contraterrene," and call ordinary bodies "terrene."

Guided by symmetry arguments, we shall assume that the contraterrene neutron, unlike the ordinary neutron, has parallel spin and magnetic moments and that the contraterrene proton and the ordinary proton are symmetric to the same extent that the positron and the electron are. The spectra of an ordinary atom and of the corresponding contraterrene atom would then be identical, and the possibility of direct spectroscopic identification of contraterrene matter would be ruled out. At this point, however, symmetry considerations suggest a further hypothesis, namely, that a fundamental terrene particle and the corresponding contraterrene particle destroy each other on collision, just as an electron and a positron are known

[1] P. A. M. Dirac, *Die moderne Atomtheorie* (by W. Heisenberg, E. Schrödinger, and P. A. M. Dirac), p. 45, Leipzig: Hirzel, 1934; G. Gamow, *Structure of Atomic Nuclei and Nuclear Transformations*, p. 14, Oxford, 1937; V. Rojansky, *Phys. Rev.*, **48**, 108, 1935.

to do. This hypothesis introduces the possibility of detecting contraterrene matter if it exists, provided that it can be observed in the process of colliding with terrene matter. The radiant energy, given by Einstein's equation $E = mc^2$, released in an electron-positron annihilation is shared by two photons, each of about half a million electron volts. In a proton-proton (one terrene, the other contraterrene) annihilation the energy per photon would be about a billion electron volts. A neutron-neutron annihilation would either yield directly a pair of billion-volt photons or reduce to a proton-proton and a positron-electron annihilation. Since electrons attract positrons and ordinary atomic nuclei would attract contraterrene nuclei, the mutual annihilation of terrene and contraterrene bodies in contact would be rapid.

We now turn to possibilities of collisions between terrene and contraterrene bodies and begin with the earth. If contraterrene matter exists in some parts of space, then some of the meteors that have their origin outside the solar system might be contraterrene. A contraterrene meteor that does not escape after entering the atmosphere would become completely converted into radiation; in particular, if it should reach the ground, it would be annihilated explosively; hence we must conclude that if contraterrene meteors exist at all, those that are large enough to penetrate deeply into the atmosphere are extremely rare.[2] Incidentally, the capture of contraterrene meteoric matter by the earth at the rate of 1 gm per square mile per year would result in the release of about 0.003 gm cal per square centimeter per minute.

The effects of a contraterrene meteor in the upper atmosphere would differ from those of a terrene meteor for two main reasons: the high penetrating power of the primary products of annihilation and the large quantity of energy released per unit mass. In particular, a contraterrene meteor would display ionizing power which is very great compared to that of a terrene meteor of the same mass and speed, and some ions would appear far from the path of such a

[2] If we disregard the effects of heating and make the rough assumption that a contraterrene meteor annihilates all the air molecules which it would strike if it did not affect their motion, we find, for example, that unless the initial top-to-bottom dimension of a contraterrene iron meteor, falling vertically without tumbling, exceeds 130 (i.e., $76 \times 13.6/7.9$) cm, the meteor would be entirely radiated away before reaching sea-level.

meteor sooner than they could arrive there by diffusion. Therefore contraterrene meteors, if they exist, can probably be detected by experiments combining ionization measurements over large portions of the atmosphere, like radio signal observations, with measurements, made in the upper atmosphere, of the energies of the annihilation products. Since the energy of the primary photons would be dissipated in the lower atmosphere through mild showers, it is not likely that the release of these photons could be ascertained on the ground by means of a single detector of a cosmic-ray type. Without a detailed study it is of course not possible to decide whether a contraterrene meteor could be identified by visual observation.

Next we shall consider a hypothetical contraterrene body revolving about the sun. Were it moving in a void, it would behave in a very ordinary way; but since it would in fact be bombarded by terrene meteoric material, it would undergo annihilation, transmutation, and mechanical disintegration. Annihilation follows from our fundamental premises. Transmutation is a concomitant of annihilation; it is certain to occur, for example, when a terrene atomic nucleus of atomic number A is annihilated in collisions with contraterrene nuclei of atomic number B which is not an integral submultiple of A, and it is likely to occur quite generally, since nuclear annihilation would consist of individual proton-proton and neutron-neutron annihilations. Mechanical disintegration, in addition to structural changes that might be caused by the sun through tides and heating, is to be expected in view of the large amounts of energy released locally by the annihilation of the larger meteoric fragments. Because of transmutation the chemical composition of the body would be gradually shifting toward the lighter elements; in particular, volatile compounds might thus be continually produced and released.

If an object having these characteristics should be observed from the earth, it would probably be classified as a comet; and the question, therefore, arises whether the comets, or at least some of them, may not actually be contraterrene bodies. The hypothesis of the contraterrene structure of comets implies the idea that, although the comets are now genuine members of the solar system, they had their origin outside the system, an idea that was first suggested for

quite different reasons and is now well known. The present hypothesis, however, does not require that, when the comets were first acquired by the solar system, they had, apart from being contraterrene, any unusual chemical composition or mechanical structure. On the basis of this hypothesis, meteors of cometary origin become of special interest; the existence of a meteorite that is definitely known to be of cometary origin would of course prove that the parent comet was terrene.

The following computation, suggested by Professor John A. Wheeler, is of interest. If a contraterrene body, bombarded uniformly by terrene matter but not appreciably heated by the sun, annihilates all this matter and absorbs half of the annihilation energy, its equilibrium temperature, computed in round numbers from the Stefan-Boltzmann law without regard for transmutation and mechanical effects, turns out to be

$$T = 1800\sqrt[4]{\mu}\,,$$

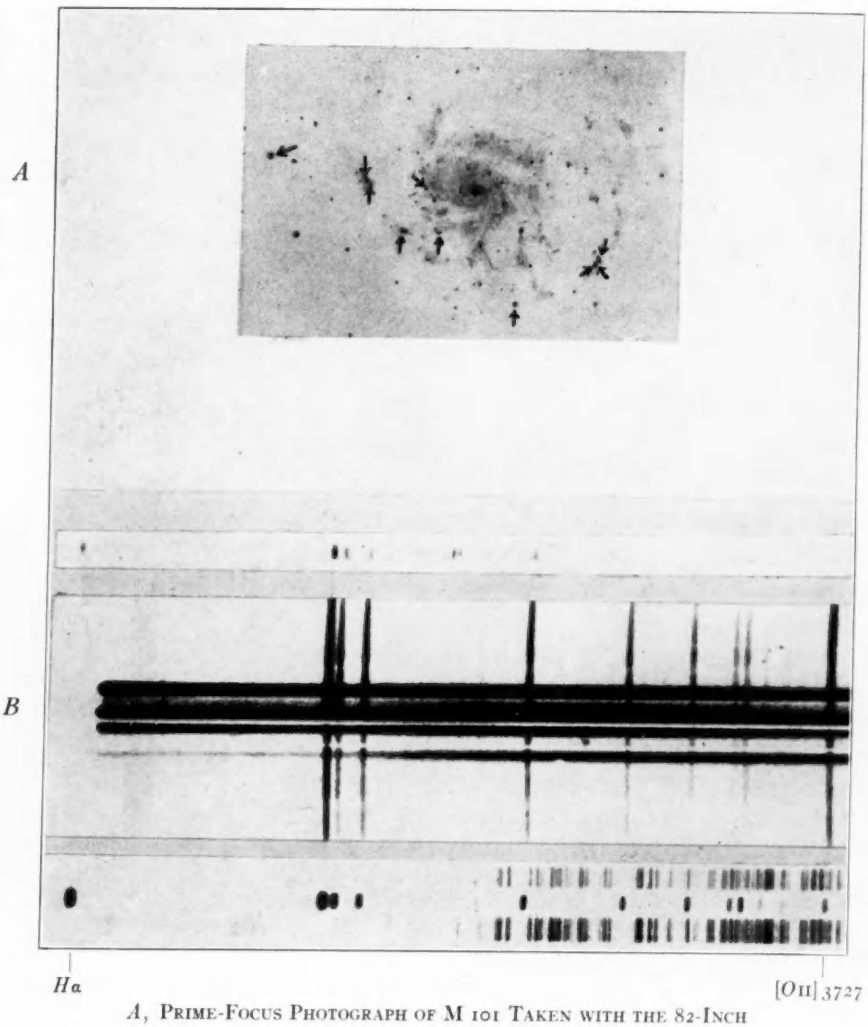
where μ is the rate, expressed in grams per square mile per minute, at which terrene matter impinges upon the body. To illustrate, we shall make two estimates of the equilibrium temperature of a contraterrene body, not heated by the sun, bombarded by terrene matter at the same rate per unit surface as that at which the earth is bombarded by meteors. Nininger[3] gives the rate of capture of meteors by the earth as 50,000 tons per day; on this basis μ is about 2×10^{-1} and T is about 1200° absolute. Wylie[4] gives the rate as 10 gm per square mile per year; on this basis μ is about 2×10^{-5} and T is about 120° absolute.

The idea of contraterrene matter suggests implications on several levels of the astronomical scale. We shall, however, stop with the few phenomena that seem most likely to provide an early proof either of the existence of this matter or of its nonexistence within the solar system.

UNION COLLEGE
SCHENECTADY, N.Y.

[3] *Our Stone-pelted Planet*, p. 91, Boston and New York: Houghton Mifflin Co., 1933.
[4] *Phys. Rev.*, **47**, 192, 1935.

PLATE II



A, PRIME-FOCUS PHOTOGRAPH OF M 101 TAKEN WITH THE 82-INCH McDONALD REFLECTOR

Positions of the emission nebulae are indicated by arrows. Ten-minute exposure on Cramer Hi-Speed plate. Contact print. Scale, 25″.6 per mm.

B, ENLARGEMENT FROM THREE FILMS TAKEN WITH THE CASSEGRAIN SLIT SPECTROGRAPH OF THE 82-INCH REFLECTOR SHOWING (*a*) OBJECT NGC 5471 IN M 101 (EXPOSURE 90 MINUTES), (*b*) ORION NEBULA (EXPOSURE 5 MINUTES), (*c*) PLANETARY NEBULA NGC 6543 (EXPOSURE 2 MINUTES)

Agfa Super Pan Press film used for all three spectrograms. Enlargement 17×. Dispersion on original spectrograms is 345 A per millimeter at $H\gamma$.

EMISSION NEBULAE IN MESSIER 101*

CARL K. SEYFERT

ABSTRACT

Ten emission objects in the spiral nebula M 101 have been found on slitless spectrograms taken with the 82-inch telescope of the McDonald Observatory. Attention is directed to one of these emission objects in particular, a small group of stars which appears to have an emission spectrum of an unusually high degree of excitation.

Patches of emission nebulosity have been found in several extragalactic nebulae, principally by Edwin Hubble,[1] by N. U. Mayall and L. H. Aller,[2] and by H. W. Babcock.[3] A 10-minute exposure with the slitless spectrograph of the 82-inch reflector of the McDonald Observatory revealed ten emission objects in the large late-type spiral M 101 (NGC 5457; $14^h1^m.4$, $+54°35'$ [1950]). The positions of the emission patches are indicated by arrows on Plate II, *A*. In the table are listed: (1) the current number; (2) the NGC designation (the most prominent condensations of M 101 have NGC numbers); (3 and 4) the rectangular co-ordinates of the emission objects, referred to the nucleus of the spiral; and (5) the emission lines observed in the order of their apparent strength on the film (Agfa Super Pan Press).

Current No.	NGC No.	ΔX	ΔY	Emission Lines
1	5447	−391″	−262″	*Hα*
2		−378	−277	*Hα*
3		−369	−284	*Hα*
4	5455	−103	−387	*Hα*, N_1, 3727 (*O* II)
5	.....	+126	−134	*Hα*
6	.....	+164	+ 17	*Hα*
7	5461	+250	−113	*Hα*, N_1, 3727, *Hβ*, *Hγ*, N_2, *Hδ*, *Hε*
8	5462	+346	+ 50	*Hα*
9		+359	+ 68	*Hα*
10	5471	+670	+170	*Hα*, N_1, N_2, *Hβ*, *Hγ*, 3727

* *Contributions from the McDonald Observatory, University of Texas*, No. 17.

[1] *Mt. W. Contr.*, No. 304, 1925; No. 310, 1926; *Ap. J.*, **62**, 409, 1925; **63**, 236, 1926.

[2] *Pub. A.S.P.*, **51**, 112, 1939.

[3] Reported at the A.A.A.S. meeting in Alpine, Texas, in May, 1939.

Hα is, without exception, the strongest feature of each of the emission spectra; in fact, it is the only line visible for all but the three brightest condensations. Among these latter objects, the *apparent* relative strengths of the lines on the film for NGC 5455 and 5461 are characteristic of galactic emission nebulae as typified by that in Orion. *Hα* is very much stronger than N_1; the *O* II doublet λ 3727 is quite strong; and for NGC 5461 in which the lines are visible, *Hβ* is stronger than N_2. These apparent relative intensities are, of course, only rough indications of the true relative intensities, since the apparent values are strongly influenced by the type of instrument and emulsion employed.

On the other hand, in NGC 5471, N_1 appears almost as strong as *Hα*; N_2 is definitely present; but only traces of *Hβ*, *Hγ*, and λ 3727 could be detected. Hence, it is probable that the degree of excitation of this object (NGC 5471) is considerably greater than that of the other condensations in emission which have been found in M 101. The high excitation of NGC 5471 is further illustrated by Plate II, *B*, which shows a slit spectrogram of this object, together with spectra of the Orion nebula and the planetary nebula NGC 6543 for purposes of comparison. All these spectra were taken on the same type of emulsion (Agfa Super Pan Press), with the same instrument (Cassegrain spectrograph of the 82-inch reflector; *f*/1 camera). It will be seen that the spectrum of NGC 5471 is not unlike that of the planetary nebula. Relative to the Balmer lines, the forbidden *O* III lines (N_1 and N_2) and the *Ne* III line at λ 3869 are considerably stronger in NGC 5471 than the corresponding lines in the Orion nebula. Using, as an excitation criterion, the ratio $N_2/3727$ (i.e., *O* III/*O* II), we see from Plate II, *B* that this ratio is approximately one for NGC 5471, whereas it is considerably less than one for the Orion nebula and greater than one for the planetary. Hence, the degree of excitation of NGC 5471 is intermediate between that of the diffuse nebula in Orion and the planetary nebula NGC 6543.

NGC 5471 is well separated from the main body of the spiral; and the observed radial velocity, as determined from the slit spectrogram (+500 km/sec), is somewhat greater than that of the spiral as determined by Humason[4] at Mount Wilson (+300 km/sec). In

[4] *Handb. d. Ap.*, 7, 555, 1936.

spite of this, it is probable that NGC 5471 is a part of M 101, since the radial velocity of the former is far too large for it to be a member of our own galaxy and too small for it to be a separate group of extragalactic nebulae showing emission lines.

On the prime-focus photograph NGC 5471 looks like a close group of five or six objects of semistellar appearance with no appreciable nebulosity involved. The other patches of emission, on the other hand, are definitely nebulous in appearance and are all within the main body of the spiral. NGC 5471 has a diameter of 15″, or about 40–50 parsecs, as against a diameter of 10–25 parsecs for the nebulous emission knots. At least two of the objects comprising the cluster NGC 5471 appear in emission on the slitless spectrogram, together with a faint continuous background, which can be seen on the original film.

McDonald Observatory
September 15, 1939

NOTES

A METEOR SPECTRUM OF HIGH EXCITATION

On the morning of October 20, 1939, in the course of an exposure made with the 10-inch Cooke prismatic camera (f = 1:4.5), a very bright meteor appeared in the field. The meteor was not observed visually, but its velocity must have been high, as a very short flash of light was noticed inside the building. A bright train was observed about 10° long; at first it was straight, but subsequently it curved ir-

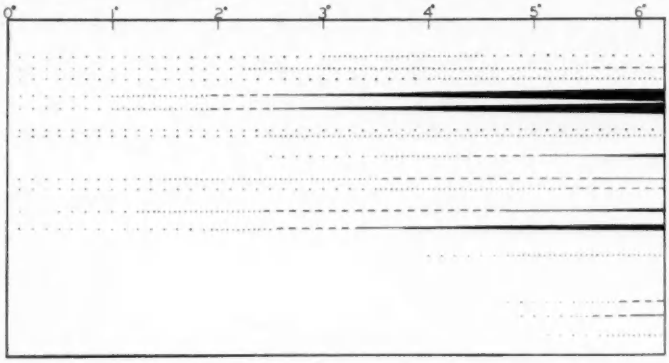


FIG. 1

regularly; it drifted backward about 10° and faded away after 5 minutes of total duration.

On the plate the spectrum first appears (Pl. III) at $\alpha = 3^h56^m$, $\delta = +42°.9$, where it is barely visible. The intensity increases continuously until the edge of the plate at $\alpha = 3^h24^m$, $\delta = +46°.0$ (1939.0).[1] As the direction of the meteor path traced backward passes close to the radiant of the Orionids, the meteor may be safely identified with this shower. It appeared at 4^h7^m local sidereal time, or 7^h30^m, G.M.T.

The spectra on the plate are short, the dispersion being 350 A/mm. The positions of sixteen lines were measured on the Gaertner machine, and their intensities were estimated at five different points.

[1] At this point the apparent magnitude was about − 7.5, as determined from trails of Jupiter.

PLATE III

METEOR SPECTRUM, OCTOBER 20, 1939 (ORIGINAL SIZE)

The measured wave lengths are given in the table and the intensities are represented schematically in the figure.

Line	λ Meteor	Element	λ Laboratory	Remarks
1	3820	*Fe* *Fe*	3820, 3826 3815	
2	3850	*Fe* *Mg*	3860 3838, 3832, 3829	Considerably out of focus
3	3880	*Fe*	3879	
4	3930	Ca^+	3934	Slightly out of focus
5	3970	Ca^+	3968	
6	4040	*Fe* *Mn*:	4046 4031, 4033, 4034	Faint, fuzzy, in the fog of lines 4 and 5 at intense end
7	4070	*Fe*	4064, 4072	
8	4140	Si^+	4131, 4128	Very sharp all through
9	4240	*Ca* *Fe* Fe^+:	4227 4236, 4250 4233	Fairly sharp, but blended
10	4280	*Cr*: *Fe*:	4254, 4275, 4289 4251, 4272, 4308, 4326	
11	4390	*Fe*	4383, 4405	Fairly sharp at the beginning; broad and fuzzy at the end
12	4490	Mg^+ Fe^+:	4481 4508	Sharp at the beginning; possibly double at the end
13	4650	Fe^+: *Cr*:	4629 4646, 4651, 4652	Faint and fuzzy
14	4950	*Fe* Fe^+:	4958, 4920, 4891 4924	Sharp
15	5060	Si^+ Fe^+:	5057, 5041 5018	Fuzzy
16	5200	*Fe* *Mg* *Cr*: Fe^+:	5192, 5233 5183, 5173, 5167 5208, 5206, 5205 5169	Fuzzy

It should be noted that some lines, such as 6, 7, and 10, appear from the very beginning, yet they never become very intense; others,

such as 4, 5, 8, and 12, while faint or invisible at the start, are very intense at the end. Evidently the lines of the first group are of low-excitation potential, while those of the second group have distinctly higher excitation potentials. In fact, the lines at the intense end are very similar in position and relative intensity to those in the spectrum of an A2 star, with the exception of the hydrogen lines, which do not, of course, appear in the meteor spectrum. This similarity has been used as a guide in the identifications which are given in the table. Thus, the lines of the ionized elements all show a rapid increase in intensity compared with the behavior of the neutral elements. Compare, for instance, line 12 of Mg^+ with 10 of *Fe*, or, again line 8 of Si^+ with 9 of *Ca*. The unusually high excitation of this spectrum is doubtless due to the high velocity of the meteor, as is evidenced by the almost instantaneous flash and by the appearance and long duration of the train.[2]

A. N. Vyssotsky

Leander McCormick Observatory
University of Virginia
October 25, 1939

NOTE ON THE SURFACE TEMPERATURE OF VENUS

Recent measurements of the total radiation emitted by gaseous carbon dioxide afford an interesting application to the determination of the surface temperature of Venus, by way of establishing an upper limit to the greenhouse effect produced by the carbon dioxide contained in this planet's atmosphere. Spectroscopic observations have so far failed to reveal the presence of water vapor or any other gases which are opaque in the infrared. By independent methods T. Dunham[1] and A. Adel[2] have estimated the amount of CO_2 in the atmosphere of Venus to be equivalent to 400 and 200 m, respectively, under standard conditions of temperature and pressure ($T = 273^\circ$ K, $p = 1$ atm). The uncertainty of these estimates is irrelevant to the argument to be given presently, as so large a mass of CO_2 has an optical depth practically infinite in the ranges of wave lengths

[2] Cf. with the conclusions of P. Millman, *Harvard Ann.*, **82**, 173–174, 1935.

[1] *Carnegie Institution of Washington Year-Book*, No. 31, 154, 1932.

[2] *Ap. J.*, **85**, 345, 1937.

covered by the emission bands and is completely transparent between these regions.

The total emission of carbon dioxide has been studied by T. Dreisch[3] and E. Eckert.[4] The work of the second author is quite comprehensive, his observations covering the range of temperatures between the boiling-point of water and 1200° C. The thickness of the radiating layer was varied between 2 cm and 3 m, and the carbon dioxide was diluted by varying amounts of nitrogen, while the total pressure was kept constant at 1 atm. Thus it was proved that the total emission of CO_2 obeys Beer's law, i.e., the total emission is determined solely by the amount of CO_2 per square centimeter, which, incidentally, is not the case for water-vapor radiation according to Eckert's findings. The validity of Beer's law for CO_2 is a highly fortunate result, as it permits us to disregard the other hypothetical constituents of the atmosphere of Venus, such as nitrogen and the noble gases. Eckert has published his measurements of the total radiation E_{CO_2} as fractions ϵ of the total radiation of a black surface of equal temperature, so ϵ is defined by

$$\epsilon = \frac{E_{CO_2}}{\sigma T^4}.$$

While ϵ increases with the thickness of the radiating layer, expressed in meter-atmospheres of CO_2, it is practically independent of T for layers thicker than about 0.10 m-atm. In other words, a thick layer of CO_2 behaves like a gray radiator with an absorption coefficient ϵ, which is independent of the temperature. Eckert's observations extend up to 2 m-atm, and their extrapolation to 200–400 m-atm leads to the value $\epsilon = 0.35 \pm 0.05$. These limits represent, in the writer's opinion, the uncertainty of both the extrapolation and the actual amount of CO_2 present on Venus.

On a slowly rotating black planet the subsolar point reaches the maximum temperature $T = 392°/\sqrt{R}$, at a distance of R astronomical units from the sun. If the planet has a spherical albedo, A, for the incoming solar radiation, while behaving like a perfect radiator in the infrared, the maximum temperature of the subsolar point

[3] *Zs. f. Phys.*, **79**, 711, 1932.

[4] VDI, *Forschungsheft No. 387*, Beilage zu *Forschung auf dem Gebiete des Ingenieurwesens*, Berlin: VDI Verlag, 1937.

is reduced to $T = 392\sqrt[4]{(1 - A)/R^2}$, or $T = 366°$ K for Venus ($A = 0.60$),[5] by neglecting the effect of the planet's atmosphere. Actually, the atmospheric CO_2 will intercept a certain part of the radiation emitted by the surface, which, according to Eckert's researches, is equal to the fraction ϵ characterizing the absorption coefficient of a thick layer of CO_2. The complemental fraction $1 - \epsilon$ can escape directly into space between the CO_2 bands. The raising of the surface temperature, resulting from the re-emission of the radiation intercepted by the atmosphere, would be most difficult to evaluate rigorously. However, an upper limit to the surface temperature follows in a simple manner from the energy balance of the total radiation, which reads

$$(1 - \epsilon)T_S^4 + \epsilon T_A^4 = 366^4 .$$

On the left-hand side the first term is the radiation escaping between the CO_2 bands from the surface (temperature T_S) and the second term is the emission of the atmosphere directed outward, which can be ascribed to a layer having the effective temperature T_A. This outgoing radiation is balanced by the insolation, producing on the atmosphereless planet a temperature of 366° K, as shown before. For obvious physical reasons, T_A must be smaller than T_S. Hence, an upper limit for T_S can be established by putting the term ϵT_A^4 equal to zero. Physically this would be equivalent to the assumption that the atmospheric CO_2 acts like a perfect reflector for its characteristic radiation and that consequently the surface behaves like a gray radiator with the absorption coefficient $(1 - \epsilon)$ for infrared radiation. With $T_A^4 = 0$ and $(1 - \epsilon) = 0.65 \pm 0.05$, T_S is found to be $408° \pm 8°$ K. The result of this admittedly quite crude procedure is nevertheless of interest, as it indicates that the atmosphere of Venus cannot raise the temperature of the surface by more than 50° at most. The actual surface temperature of the subsolar point will be somewhere between 366° and 408° K, probably closer to the second value. In any case, it appears to be higher than the terrestrial boiling-point of water.

RUPERT WILDT

PRINCETON UNIVERSITY OBSERVATORY
November 1939

[5] H. N. Russell, *Ap. J.*, **43**, 173, 1916.

NEW WHITE DWARFS, SUBDWARFS AND BINARY STARS*

This note is a continuation of one published ten months ago.[1] Only some of the more striking objects are given here. A complete list of the stars observed will follow later.

Table 1 contains six new white dwarfs. Although no trigonometric parallaxes are as yet available for these stars, their white-dwarf nature is established from the spectra in four cases. For the two F stars the possibility of the stars being subdwarfs cannot be entirely excluded without more data.

Next follow seventy-nine probable subdwarfs. Most of the objects for which spectra were published earlier are included here, for the sake of uniformity, with the writer's spectral types. The present list, together with the preceding one,[1] forms a fairly complete list of the known subdwarfs with annual proper motion in excess of $0''.30$. Not included in this list are the many objects for which the subdwarf nature is based only on the spectral characteristics. Our list is selected on the basis of two criteria: transverse motion[1] and reliable trigonometric parallax.

It is possible that a few objects now classed as subdwarfs do not belong to that class because the proper motion found in the blink microscope is considerably too large, or perhaps even spurious. Such cases may be eliminated by a redetermination of the proper motion or by a closer study of the spectrum, which in that case should not show the subdwarf features.[2] In this manner the spurious proper motion of $+54^{\circ}2461$, $10^{m}.2$ Ao, was detected. Cases of this kind will, however, be quite rare.

The assignment of some faint G stars to the subdwarfs rather than to the white dwarfs is tentative and is based on the presence of the G band in the spectrum, which seems to be absent in white dwarfs.

The present paper contains the first M-type subdwarfs definitely established. All three stars appear to be about 2 mag. below the

* *Contributions from the McDonald Observatory, University of Texas*, No. 18.

[1] *Ap. J.*, **89**, 548, 1939.

[2] For a discussion of these features, as well as of other properties and their interpretation, cf. *Ann. les d'astrophysique* (now in press).

TABLE 1

Star	α	δ	m_{vis}	μ	Sp	π (Sp)	π (tr.)
Wolf 1	0:08.5	− 0:14	15.0	0″.53	Ao	.....	.....
Wolf 1516	1 12.7	+15 40	13.6	0.69	Con	.....	.....
Ross 548	1 31.2	−11 51	13.5	0.42	Ao	.....	.....
Ross 22	3 57.9	+48 56	14.1	0.77	F	.....	.....
Ross 640	16 24.9	+36 58	13.2	0.89	F	.....	.....
Wolf 672A	17 13.6	+ 2 04	14.4	0.56	Wh.*	.....	.....
+71°31	0 37.2	+71 38	10.2	0.35	Fo	0″.003	0″.004±9
Wolf 1504	0 39.1	+29 22	13.0	0.49	G5	.003	.....
Wolf 56	1 01.9	+62 59	11.5	1.04	K2+	.010	.....
Wolf 79	1 32.6	+ 9 49	12.0	0.40	Go	.003	.....
Wolf 1065	1 37.4	+16 38	12.3	0.48	Go:	.003	.....
+72°94	1 38.9	+72 58	10.1	0.30	Fo	.004	.....
Wolf 107	1 55.7	+ 3 33	11.7	0.42	G2	.004	.....
Wolf 110	1 57.4	+ 5 14	11.9	2.41	K4	.012	.038±6
Wolf 125	2 08.2	+15 31	13.4	1.08	K1	.004	− .006±7
Ross 18	2 09.2	+31 56	11.8	0.48	G2	.004	.....
Ross 557	2 38.4	+27 15	13.7	0.54	G5	.002	.....
+33°529	2 45.9	+34 00	9.7	1.39	K3+	.029	.066±5
Wolf 134	3 07.6	+18 29	14.0	1.74	K3.	.004	.050±8
Ross 348	3 10.1	+51 56	13.4	0.41	G	.002	.....
Ross 373	3 12.0	+23 16	14.0	0.70	G5	.002	.....
Ross 374	3 21.1	+23 26	10.8	0.43	F3	.003	.....
+66°268	3 21.8	+66 25	9.7	1.56	G5	.013	.023±7
Ross 578	3 33.4	−11 45	13.1	3.16	M2	.025	.060±6
Ross 581	3 36.2	− 8 09	12.8	0.35	G8	.004	.....
Ross 31	4 26.5	+50 23	13.6	0.44	K2	.004	.....
Wolf 1540	4 49.0	+ 6 54	13.0	0.35	F8	.002	.....
Ross 385	4 56.6	+15 24	13.8	0.35	G2	.002	.....
Kapteyn's star	5 07.7	−44 59	8.8	8.78	Mo	.115	.262±6
Ross 65	5 16.6	+33 06	12.0	0.74	K1+	.007	.....
+19°1185A	5 57.3	+19 23	9.2	0.93	G1	.013	.028±6
+19°1185B	5 57.3	+19 23	12.8:	0.93	K3	.006:	.028±6
+31°1684	7 47.2	+30 55	8.3	1.96	G2	.020	.036±5
Ross 618	8 04.5	+10 28	13.6	0.40	G5	.002	.....
+54°1216	8 11.7	+54 25	9.5	0.64	F2	.006	.008±9
−15°2546	8 36.2	−15 59	9.6	0.60	F2	.005	.025±6
−12°2669	8 42.0	−13 00	10.0	0.38	A2	.002	− .015±11
Ross 683	8 45.1	+ 8 00	11.4	0.62	G5	.005	.....
−3°2525	8 54.1	− 3 37	9.8	0.76	F5	.006	.026±12
Ross 885	9 08.6	+20 27	13.4	0.46	K1	.004	.....
+1°2341	9 35.6	+ 1 29	10.6	0.54	Fo	.003	.....
+44°1910	9 43.2	+44 46	10.9	0.21	A5	.002	.....
+14°2151	9 43.5	+14 14	8.4	0.83	Fo	.008	.026±4
Wolf 336	9 55.5	+33 21	12.7	0.38	G8	.004	.....
Wolf 338	9 56.2	+35 09	15.6	0.32	K3:	.002	.....
Ross 891	10 21.7	− 1 51	13.2	0.34	G2	.002	.....
+29°2091	10 42.0	+28 56	10.4	1.0	F8	.006	.016±9
Wolf 365	11 05.9	+ 6 59	10.8	0.82	G2	.006	0.022±6
(L1259–59)	11 12.4	+17 48	13.7	0.88	G5	0.002	†

* Provisional; a spectrum in the visual region shows the star to be white; from the companion (type M3) we find the absolute magnitude to be about 11.

† Identification uncertain.

TABLE 1—*Continued*

Star	α	δ	m_{vis}	μ	Sp	π (Sp)	π (tr.)
ξ UMa(M)	11:12.9	+32:06	4.6	0".73	F9	0".09	0".138±6
Ross 111	11 25.8	+59 45	13.0	0.68	K3+	.006	.....
AC77°4245	11 26	+77 12	11.3	0.60	Go	.004	.....
Ross 451	11 34.7	+67 53	12.3	3.20	K4	.010	.041±7
Ross 452	11 54.2	+68 21	12.2	0.50	G4	.004	.....
Ross 636	12 01.1	+10 46	12.4	0.38	Go	.003	.....
Wolf 1440	12 12.2	+21 37	12.1	0.74	G5	.004	.....
Wolf 458	12 55.2	+ 6 07	14.5	(0.6)	Go	.001	‡
+10°2519A	13 06.4	+10 09	8.6	0.60	G2	.018	.051±6
+10°2519B	13 06.4	+10 09	12.3	0.60	K5	.014	.051±6
+34°2476	13 54.8	+34 23	10.3	0.54	Fo	.003	.023±7
+6°2932	14 38.4	+ 6 15	10.4	0.91	G3	.009	− .046±9
−21°4009	14 54.2	−21 36	8.5	0.77	F2	.009	.030±9
−15°4042	15 04.7	−15 54	9.1	3.68	G8	.020	.040±4
−15°4041	15 04.7	−15 59	9.4	3.68	Ko	.020	.040±4
−10°4149	15 37.7	−10 36	7.2	1.18	F2	.016	.033±6
+42°2667	15 59.9	+42 32	10.0	0.51	F4	.005	.....
L1130–91	16 07.3	+ 5 47	12.1	0.69	K2	.007	.....
Ross 530	16 15.6	+22 53	11.4	0.47	G2	.005	.....
+13°3683	18 28.7	+13 03	10.6	0.31	F3	.003	− .002±10
+35°3659sf	19 27.6	+35 57	10.6	0.56	F2	.003	.007±7
L–Ebb. 21	20 05.8	+56 52	13.4	0.50	K2	.004	.....
Wolf 1106	21 05.6	+59 22	13.0	2.14	Mo	.017	.037±8
Wolf 923	21 28.5	− 7 17	15.1	0.51	K2:	.002	.....
Her. 39	21 50.8	+32 10	11.0	0.75	G2	.006	− .024±11
+17°4708	22 06.7	+17 36	9.3	0.54	F2	.006	.....
Ross 272	22 06.8	+17 34	11.9	0.37	G2	.004	.....
Wolf 1216	22 16.2	+29 17	12.3	0.75	K3:	.008	.....
Ross 666	22 22.6	+54 02	12.2	0.42	A2	.001	.....
Wolf 1037	22 23.7	+ 5 20	14.0	1.57	K2:	.003	.052±8
Ross 288	22 39.7	− 2 52	11.4	0.85	G8	.007	.....
−0°4470	23 04.4	+ 0 12	10.2	1.32	G2	.009	.....
+38°4955	23 08.9	+38 53	11.0	0.57	F8	.004	.....
−14°6437AB	23 11.9	−14 22	8.3	1.37	F8	.011	.019±7
+59°2723	23 21.9	+60 04	10.4	0.46	F4	.004	.019±9
−42°16457	23 41.2	−42 07	7.5	0.88	A3	0.008	0.026±7

‡ Wolf found μ = 0".61; Luyten states that μ is smaller but does not give its value.

average main sequence. It is of interest to note that Ross 578 is shown also spectroscopically to be different from ordinary M dwarfs.

The most extreme subdwarfs thus far found are Wolf 134 and Wolf 1037. From the proper motion we find that they must be more than 3.5 mag. below the main sequence; the trigonometric parallaxes suggest that the deviation may be as much as about 6 mag. Additional parallax determinations will be of interest, particularly for these two stars.

There are twenty-six entries in Table 1 for which independent

trigonometric parallaxes are available and for which $\pi(\text{Sp}) \leqslant 0.01\mu$. These stars would have been classified as subdwarfs even if the trigonometric parallax had not been known. It is of interest to obtain from these stars the most probable ratio of $\pi/\pi(\text{Sp})$. Taking into account the greatly different weights of the twenty-six determinations, we find for the weighted mean value 2.2 ± 0.2 (p.e.). This

TABLE 2

Star	α	δ	μ	π (tr.)	m_V	Sp.	π (Sp.)	d
−13°249A	1:18.0	−13:29	0.46	0".049 ± 10	7.5	G8	0".042	.....
−13°249B	.....	.....	.....	.....	10.5	dK6	.042	40".±
+3°275A	1 56.9	+ 3 28	.49	.....	10.0	K3+	.025	.....
+3°275B	.....	.....	.....	.....	13.0:	M2	.026:	15.
L-Ebb. 7A	2 30.8	+31 37	.40	.....	13.5	M2	.021	.....
L-Ebb. 7B	.....	.....	.....	.....	14.3	.....	.....	3.
Ross 45A	5 28.7	+10 16	.36	.....	12.4	M5	.096	.....
Ross 45B	.....	.....	.....	.....	14.4	M6	.060:	5.
−3°2001	7 35.0	− 3 22	.29	.090 ± 7	7.2	K2+	.074	.....
−3°2002	.....	.....	.....	.....	8.7	dK5	.072	50.
Ross 825A	21 07.1	+33 07	.48	.....	11.9	K3+	.010	1½
Ross 825B	.....	.....	.....	.....	13.1			
Ross 773A	21 12.8	+20 28	.55	.....	12.2	M3	.05	.....
Ross 773B	.....	.....	.....	.....	13.6	M4	.04	3.2
Ross 223A	22 41.2	+44 02	0.34	0.033 ± 9	10.9	K4	0.021	.....
Ross 223B	.....	.....	.....	.....	12.9	.....	.....	2.

means that the harmonic mean value of the transverse motion for these stars is 215 km/sec. This mean ratio corresponds to a deviation of 1.7 ± 0.2 mag. from the average main sequence. The range seems to be from about 1 mag. (a somewhat arbitrary limit) to perhaps 5 mag. below the main sequence.

Table 2 contains some new binaries found recently. For the two wide pairs proper-motion data will be required to establish the physical relationship definitely.

G. P. KUIPER

McDONALD AND YERKES OBSERVATORIES
January 1940